

2

AD-A255 748

3



Demodulation Processes In Auditory Perception

Lawrence L. Feth
Division of Speech and Hearing Science

DTIC
ELECTE
SEP 14 1992
S B D

Air Force Office of Scientific Research
Bolling Air Force Base, D.C. 20332-6448

Grant No. AFOSR-89-0227
Final Report

92-25074



19/88

August 1992

267200

92 9 11 016

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

12 13



Demodulation Processes In Auditory Perception

Lawrence L. Feth
Division of Speech and Hearing Science

Air Force Office of Scientific Research
Bolling Air Force Base, D.C. 20332-6448

Grant No. AFOSR-89-0227
Final Report
RF Project No. 767249/721777

August 1992

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
1a. REPORT SECURITY CLASSIFICATION			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT		
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE			4. PERFORMING ORGANIZATION REPORT NUMBER(S)		
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION The Ohio State University		6b. OFFICE SYMBOL (if applicable)		7a. NAME OF MONITORING ORGANIZATION AFSR / NL	
6c. ADDRESS (City, State, and ZIP Code) 110 Pressey Hall Columbus, Ohio 43210		7b. ADDRESS (City, State, and ZIP Code) Building 410 Building AFB DC 20332-6445		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFSR-89-0227	
8a. NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR		8b. OFFICE SYMBOL (if applicable) NL		10. SOURCE OF FUNDING NUMBERS	
8c. ADDRESS (City, State, and ZIP Code) Bldg 410 Building AFB DC 20332-6445		PROGRAM ELEMENT NO C1102F		PROJECT NO 2313	
		TASK NO A6		WORK UNIT ACCESSION NO	
11. TITLE (Include Security Classification) Demodulation Processes in Auditory Perception					
12. PERSONAL AUTHOR(S) Lawrence L. Feth					
13a. TYPE OF REPORT Final		13b. TIME COVERED FROM 1-12-88 TO 31-5-92		14. DATE OF REPORT (Year, Month, Day) 92 August 15	
15. PAGE COUNT					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Complex Sound Discrimination		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This document reports the accomplishments of a project on the application of the Envelope-Weighted Average of Instantaneous Frequency (EWAIF) model to the processing of complex, time-varying sounds. We consider the task of human listeners to be one of recovering information imposed on the sound stream by a variety of sources. These include speech, music and other environmentally-important signals. Information is encoded in amplitude (envelope) and angle (frequency or phase) modulations of the sound stream "carrier". The human listener must demodulate the stream to recover the information. EWAIF first demonstrated that these modulations interact and could provide discrimination cues even for "steady-state" signals such as those used in profile analysis or co-modulation masking. This project revised the EWAIF model into the IWAIF version - Intensity (envelope-squared) weighting leads to greater computational efficiency (via the FFT) and to an intuitively appealing representation. The IWAIF calculation leads to the "center-of-gravity" of the spectrum. Tracking frequency modulations imposed on a narrow bandwidth carrier, then may be thought of as tracking the spectral center of gravity. Work continues on the extension of the IWAIF model to handle processing of signals with multiple modulation sources and to refining the short-term tracking abilities of the model.					
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION		
22a. NAME OF RESPONSIBLE INDIVIDUAL Lawrence L. Feth, Jr. 2/14/92			22b. TELEPHONE (Include Area Code) 614/292-8202/211-5221		22c. OFFICE SYMBOL NL

25 AUG 1992

FINAL TECHNICAL REPORT

1 DEC. 1988 - 31 MAY 1992

Introduction

The goal of this three-year project was to better understand the human listener's ability to process important classes of complex sounds. In particular, we adopted a modulation-demodulation (mo-dem) model of auditory perception. That is, we took the position that useful information in the sound stream reaching the human listener may be characterized as modulations of signal amplitude and angle. Angle modulation can be expressed as either phase or frequency modulation. The listener's task then, in extracting information from the sound stream, may be characterized as a demodulation process. For this project, we have focused primarily on the processing of frequency-modulated (FM) signals.

Work conducted by the PI and several of his students and colleagues, prior to the initiation of this project, had led to the formulation of the EWAIF model for complex sound discrimination. EWAIF is an acronym for envelope-weighted average of instantaneous frequency. We had found that listener performance could often be predicted by calculating the EWAIF values for complex signal pairs that were discriminable in forced-choice paradigms. We began the series of studies with common-envelope signal pairs as defined by Voelcker in his work on a Unified Theory of Modulation. Our first attempt to relax the common envelope restriction appeared to be quite successful. We applied the EWAIF model to the array of sinusoids used in early profile analysis studies, and offered an alternative explanation for the apparent enhanced sensitivity to amplitude increments embedded in profile arrays of greater signal densities. That is, Green and his associates reported that listeners could detect an ever smaller increment to the center component of the profile array as more sinusoids were added to the band occupied by the array. We calculated EWAIF values for flat versus incremented arrays and showed that a reliable pitch shift was produced by the increment. The shift is dependent on the relative amplitudes of the tones not the absolute levels, thus roving-level manipulations were ineffective in rendering the cue unusable. Further, as tone density increased and just-detectable amplitude increments decreased, the difference in EWAIF values remained approximately constant. We thus concluded that the enhancement in profile performance with increased tone density was probably related to the pitch cue rather than to profile analysis, per se. This work was reported at the Complex Sound Workshop sponsored by AFOSR at Sarasota, FL in 1986.

Recent work by Versfeld and Houtsma, however, has indicated that the common-envelope restriction cannot be disregarded with most narrow band-width signals. In two reports they have demonstrated the inadequacy of EWAIF predictions for complex signal pair discriminations. We have replicated some of their results and are currently conducting an investigation of modifications required to use a demodulation model for these results. Essentially, when the signal envelopes differ, as they do in Versfeld and Houtsma's signals, the listener is likely to use both amplitude and frequency information to distinguish between the signals. Anantharaman, Krishnamurthy and the PI are currently work-

ing on a multi-channel version of the EWAIF model that will incorporate envelope cues as well. (Preliminary results are shown in Figure 1, appended to this report.)

List of research objectives and cumulative progress

1. Step vs. Glide discrimination

Two manuscripts for publication have resulted from work on this phase of the proposal. John Madden's dissertation, which extended the original step vs. glide discrimination task to listeners with sensori-neural hearing loss, appeared in the April 1992 issue of *Journal of Speech and Hearing Research*. The manuscript detailing our earlier work with normal-hearing listeners, was submitted to *Journal of the Acoustical Society of America* in April 1992. It is currently undergoing some minor revisions requested by the reviewers. We anticipate that it will appear in the journal by late 1992 or early 1993. Details of this work can be found in copies of these manuscripts which are appended to this report.

2. Multi-channel EWAIF model

Work on extending the original EWAIF model to incorporate an approximation of the peripheral filtering of the human auditory system has been especially fruitful. Jayanth Anantharaman, a graduate student in Electrical Engineering under the supervision of Co-Investigator Ashok Krishnamurthy, tackled this phase of the project as his masters thesis project. Anantharaman first devised a computationally more efficient form of the original EWAIF model, which we have dubbed the IWAIF model. That is, Intensity-weighted average of instantaneous frequency. The model output can be computed very efficiently in the frequency domain using the FFT, rather than the time domain. Further, the result of the frequency domain calculation has an appealing interpretation in terms of the center-of-gravity of the sound spectrum. A convention presentation of the initial work was followed by a manuscript submitted to *Journal of the Acoustical Society of America* in February 1992. A transition of Associate Editors and the apparent overlap of content in a similar manuscript by Dai from Green's group at Florida has led to a delay in the processing of the manuscript. Revisions are underway for a submission of the revised manuscript by the end of August. A copy of the Anantharaman, et al. manuscript is appended to this report. Work on the extension to a multi-channel model continues as Anantharaman has begun to work on his PhD project.

3. Single-step vs. glide experiments

This series of experiments was begun in collaboration with R. Gerren at Kansas University in 1987-88. Data collection was completed before the PI moved from Kansas to Ohio State. Gerren visited Ohio State in No 1989 as a consultant on the current project, but a manuscript for publication has not been forthcoming. Lack of support for his work at Kansas has distracted Gerren from completing the work. With some help, the PI will produce the final draft of the paper for submission.

4. FM transitions with amplitude contours

The preliminary work on this portion of the project, initiated by Y. Y. Qi were presented at the Nov. 1990 meeting of the Acoustical Society in San Diego. No further work was conducted on this research line because work on other aspects of the project required more time than originally projected. These preliminary results will be of value in the implementation of a multi-channel IWAIF model for speech-like complex sounds.

5. Glide direction and slope discrimination

The completion of experiments described under this phase of the proposal required the development of software for the "real-time" generation of frequency-modulated tones. Further work was required to conduct roving-frequency discrimination experiments running three listeners in independent adaptive-tracking tasks. Chien Yeh Hsu developed the required software so that we could conduct these experiments efficiently. A manuscript describing the software was submitted to *Behavior Research Methods, Computers and Instrumentation* in April 1992. We are awaiting an editorial decision.

The initial experimental work on this phase of the project served as the masters thesis for Mary Neill. Her work was conducted prior to the completion of the adaptive tracking portion of the software described above. Preliminary reports of the work were presented at both ARO and Acoustical Society meetings. Ms. Neill has elected not to continue in the graduate program at this time, and work on preparation of a manuscript for publication has been retarded. A copy of the thesis work is appended to this report. We anticipate that a publishable manuscript will be finished in fall 1992.

Considerable time was spent in developing an adaptive tracking procedure for the determination of just-discriminable frequency-glide slope. Prior work had required blocks of trials at fixed slope differences to establish full psychometric functions. Once we determined that these psychometric functions were monotonic and reasonably well-behaved, we moved to an adaptive testing paradigm. Our initial results from the adaptive testing led us on a long chase for possible procedural or equipment artifacts, because of an apparent hysteresis in the FM glide slope discrimination thresholds. When the target slope approached the standard slope from "above" (i.e., the target was steeper) the adaptive routine settled into a threshold slope difference that could be as much as ten times larger than when the routine approached from "below" (i.e., the target was flatter). Introducing roving starting frequency conditions further complicated the results. (See Figure 2 appended to this report.)

We now are confident that these unexpected results are not simply due to artifact in the experimental procedures. Similar hysteresis has been reported by Porter, Cullen Collins and Jackson [*J. Acoust. Soc. Amer.* 90, 1298-1308, 1991]. Porter et al., were investigating formant transition onset frequencies. While these hysteresis effects appear to be "real", we have not formulated a reasonable explanation for their occurrence.

Work on this phase of the project continues with the PhD project of Hsu, who is developing a short-term running version of the IWAIF model to predict performance in FM glide slope and direction discrimination. The model will incorporate the "front-end" of the Patterson-Holdsworth "Auditory Sensation Processor" model. The series of "source-filter"

discrimination experiments described in the proposal will be conducted as part of Hsu's dissertation work.

6. Additional experiments

Several experiments not described in the 1988 proposal have been conducted as part of this project. One has been mentioned in section 1 above. John Madden's dissertation project extending the glide vs. step discrimination task to listeners with sensori-neural hearing loss was not anticipated in the proposal. In addition to being of value for understanding the deleterious effects of hearing loss on the ability of human listeners to process complex sounds, the project allows us to probe a bit further into possible physiological mechanisms underlying this ability. We now are formulating plans to extend the paradigm to persons implanted with a multi-channel cochlear implant. This work will likely be the basis for a dissertation project by Ina Bicknell. In addition to the information gained on the signal processing abilities of implant wearers, this project may allow a direct test of our notion that neural synchrony is essential for optimum performance on this task. Since we can drive auditory nerves directly through the implant processor, we should gain some insight into the underlying physiology for this task.

Another project not anticipated in the original proposal was the study of dichotic vs. diotic profile analysis conducted by Gail Wightlaw for her dissertation project. The PI has asserted that some part of the profile analysis processing was due to the frequency-modulation artifact generated when one tone of a multi-tone profile array was increased in level. Since such FM artifacts would be difficult to demonstrate in true dichotic stimuli, we designed a test of dichotic vs. diotic profile analysis. Little was reported in the profile literature on the possibility of profile analysis in dichotic listening. What was available seemed contradictory, with Green's associates claiming little support for dichotic profile analysis capabilities in their listeners, but Fantini, et al reporting reduced but substantial dichotic profile analysis results. Whitelaw's work supports that of Fantini et al. A manuscript for submission to *Journal of the Acoustical Society of America* is nearly complete. (A copy of the draft is appended to this report.)

Finally, the entrée into dichotic signals led us to the reports by Clifton and her associates on the dynamic effects of prior stimulation on echo suppression. The published work was always reported for sound field listening conditions. Since the precedence effect and other demonstrations of echo suppression can be demonstrated under headphones, we questioned the apparent lack of headphone listening data. Pat Burton chose to follow this question for her masters thesis work. The thesis is nearly complete. We anticipate a defense by Sept. 1992. A copy of the completed work will be forwarded at that time.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist.	Avail and/or Special
A-1	

DTIC QUALITY INSPECTED 3

Participating Professionals

Lawrence L. Feth, PhD	Principal Investigator
Ashok K. Krishnamurthy, PhD	Co-Investigator
Yingyong Qi, PhD	Post-doctoral Fellow (3/89 - 7/89)
Mary E. Neill, MA	Grad. Research Assoc. (to 9/90)
Chien Yeh Hsu, MS	Grad. Research Assoc. (1/90 - 6/92)
John B. Madden, PhD	Grad. Research Assoc. (1/90 - 9/90)
Joel B. Treadway, BS	Grad. Research Assoc. (9/90 - 6/91)
Gail M. Whitelaw, PhD	Grad. Research Assoc. (1/91 - 6/91)
Patricia Burton, MA	Grad. Research Assoc. (9/91 - 12/91)
Jayanth N. Anantharaman, MS	Grad. Research Assoc. (no cost)*
Ina R. Bicknell, MS	Grad. Research Assoc. (no cost)*

* support provided by non-grant funds

Publications and presentations

Auditory temporal acuity for dynamic signals. L. L. Feth, M. E. Neill and A. K. Krishnamurthy. Presentation to The Acoustical Society of America, Nov. 1989, St. Louis, MO [Abstract: *J. Acoust. Soc. Amer.*, 86, S122 (1989)]

Discrimination of frequency-glide direction. M. E. Neill and L. L. Feth. Presentation to The Acoustical Society of America, May 1990, State College, PA. [Abstract: *J. Acoust. Soc. Amer.*, 87, S23 (1990)]

Auditory temporal acuity in hearing-impaired listeners for frequency-modulated signals. J. B. Madden and L. L. Feth. Presentation to The Acoustical Society of America, May 1990, State College, PA [Abstract: *J. Acoust. Soc. Amer.*, 87, S157 (1990)]

Envelope-weighted average of instantaneous frequency (EWAIF) model applied to narrow bandwidth signals. L. L. Feth. Presentation to The Acoustical Society of America, Nov. 1990, San Diego, CA [Abstract: *J. Acoust. Soc. Amer.*, 88, S48 (1990)]

Discrimination of FM glide direction and slope. L. L. Feth, M. E. Neill and C. Y. Hsu. Presentation to The Acoustical Society of America, Nov. 1990, San Diego, CA [Abstract: *J. Acoust. Soc. Amer.*, 88, S48 (1990)]

Discrimination of frequency-modulated glides using a roving-frequency paradigm. L. L. Feth, M. E. Neill and C. Y. Hsu. Presented to The Association for Research in Otolaryngology, 14th Midwinter meeting Feb. 1991.

Temporal resolution of frequency-modulated signals by hearing-impaired listeners. J. P. Madden and L. L. Feth. Presentation to The Acoustical Society of America, May 1991, Baltimore, MD [Abstract: *J. Acoust. Soc. Amer.*, 89, S2008 (1991)]

Dichotic profile analysis, G. M. Whitelaw, C. Y. Hsu and L. L. Feth. Presentation to The Acoustical Society of America, May 1991, Baltimore, MD [Abstract: *J. Acoust. Soc. Amer.*, 89, S1912 (1991)]

Identification of Initial Stop Consonants Processed by the Patterson-Holdsworth Model, R. A. Fox and L. L. Feth. In Auditory Physiology and Perception Edited by Y. Cazals, L. Demany and K. Horner, Pergamon Press. Oxford (1992).

Temporal resolution of frequency-modulated tones, J. P. Madden and L. L. Feth. Presented at the annual meeting of the American Speech-Language-Hearing Association, Atlanta, GA. (1991). [Abstract: *asha*, p. 166, October, 1991]

Temporal Resolution in Normal and Hearing-Impaired Listeners using Frequency Modulated Stimuli. J. P. Madden and L. L. Feth. *J. Speech and Hearing Res.*, 35, 436-442, (1992).

Auditory Temporal Acuity using Frequency Modulated Sinusoids. L. L. Feth and J. P. Madden. Revision submitted to *J. Acoust. Soc. Amer.* August (1992).

Using the Ariel DSP-16 as a Signal Generator for Psychoacoustics Experiments. C. Y. Hsu and L. L. Feth, Submitted to: *Behav. Res. Meth., Comp. & Instru.* April (1992).

IWAIF: Intensity Weighted Average of Instantaneous Frequency Model for Frequency Discrimination. J. N. Anantharaman, A. K. Krishnamurthy and L. L. Feth, Submitted to: *J. Acoust. Soc. Amer.* March (1992).

Patents and Inventions

No patentable inventions have resulted from this research

General Statements

Often the impact of a research project is assessed solely by the number of publications it has produced. This project might be judged to have been of little impact if numbers of papers published (to date) were the only criteria. The PI notes here that he has been lax in getting the results of this work submitted to the journals as promptly as he should have. Several more manuscripts derived from these three years of support will be completed and submitted over the next several months. The work supported by this grant has had an impact on the field. We note here the work by Richards, Onsan and Green, "Auditory profile analysis: Potential pitch cues", [*Hearing Research*, 39, 27-36, (1989)]; Kidd, Mason, Uchanski, Brantley and Shah, "Evaluation of simple models of profile analysis using random reference spectra", [*J. Acoust. Soc. Amer.*, 90, 1340-1354, (1991)]; and Versfeld and Houtsma, "perception of Spectral Changes in Multi-tone Complexes", [*The Quarterly J. of Experimental Psychology*, 43A, 459-479, (1991)]. Modifications of the EWAIF model have been reported by B. Berg and H. Dai (currently, or formerly, of Green's research group at Florida).

Common Envelope Signals

I. Introduction

Voelker's (1966) basis signal pair $(A, f_1; A + \Delta A, f_2)$ and $(A + \Delta A, f_1; A, f_2)$ are known to have the same envelope. How do we extend this idea of common envelope pairs to multi-component signals? Two possibilities are discussed below. Employing these signals as auditory-stimuli in actual experiments is as yet unclear.

II. Type I

We claim that the following signal pair have a common envelope. This pair, $s_1(t)$ and $s_2(t)$, is derived from the basic pair by duplication at frequencies ω_i from them.

$$s_1(t) = \sum_m a \cos(\omega_a + \omega_m)t + b \cos(\omega_b + \omega_m)t \quad (1)$$

$$s_2(t) = \sum_m b \cos(\omega_a + \omega_m)t + a \cos(\omega_b + \omega_m)t \quad (2)$$

Let us calculate the envelope of signal $s_1(t)$. The corresponding analytic signal $m_1(t)$ is

$$m_1(t) = \sum_m a \cos(\omega_a + \omega_m)t + b \cos(\omega_b + \omega_m)t + j a \sin(\omega_a + \omega_m)t + j b \sin(\omega_b + \omega_m)t \quad (3)$$

$$= \sum_m a \exp j(\omega_a + \omega_m)t + b \exp j(\omega_b + \omega_m)t \quad (4)$$

$$= (a \exp j\omega_a t + b \exp j\omega_b t) \sum_m \exp j\omega_m t \quad (5)$$

The envelope, $e_1(t)$ is given by

$$e_1(t) = |m_1(t)| \quad (6)$$

$$= |a \exp j\omega_a t + b \exp j\omega_b t| \left| \sum_m \exp j\omega_m t \right| \quad (7)$$

Similarly, the envelope of signal $s_2(t)$ is

$$e_2(t) = |b \exp j\omega_a t + a \exp j\omega_b t| \left| \sum_m \exp j\omega_m t \right| \quad (8)$$

As in the case of the basic Voelker pair,

$$|a \exp j\omega_a t + b \exp j\omega_b t| = |b \exp j\omega_a t + a \exp j\omega_b t| = a^2 + 2ab \cos(\omega_b - \omega_a)t + b^2 \quad (9)$$

Therefore signals $s_1(t)$ and $s_2(t)$ have the same envelope. We shall further investigate the instantaneous frequency.

III. Type II

The following pair of signals are spectral ramps. The frequencies of the components vary linearly while their respective amplitudes vary linearly on a log scale.

$$s_1(t) = \sum_{m=0}^N A \alpha^m \cos(\omega_c + m\omega_0)t \quad (10)$$

$$s_2(t) = \sum_{m=0}^N A \alpha^{N-m} \cos(\omega_c + m\omega_0)t \quad (11)$$

The analytic signal corresponding to $s_1(t)$ is

$$m_1(t) = \sum_{m=0}^N A \alpha^m \exp j(\omega_c + m\omega_0)t \quad (12)$$

$$= A \exp j\omega_c t \sum_{m=0}^N (\alpha \exp j\omega_0 t)^m \quad (13)$$

The envelope is then

$$e_1(t) = |m_1(t)| \quad (14)$$

$$= A \left| \sum_{m=0}^N (\alpha \exp j\omega_0 t)^m \right| \quad (15)$$

Similarly the envelope of $s_2(t)$ is

$$e_2(t) = A \left| \sum_{m=0}^N \alpha^{N-m} \exp j m \omega_0 t \right| \quad (16)$$

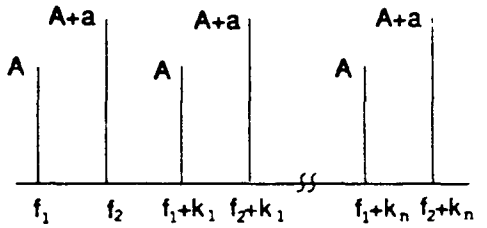
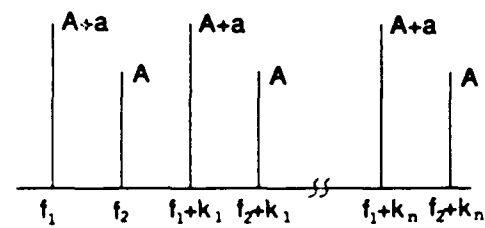
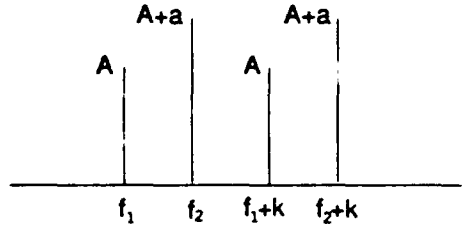
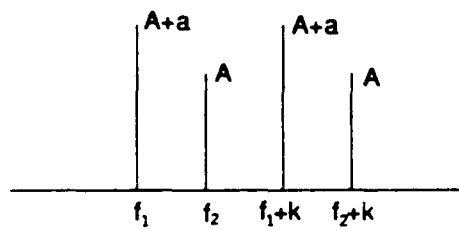
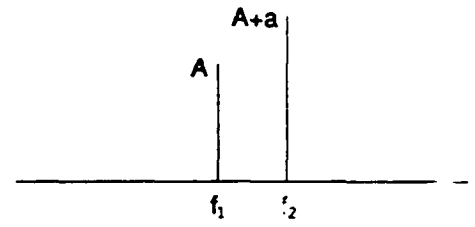
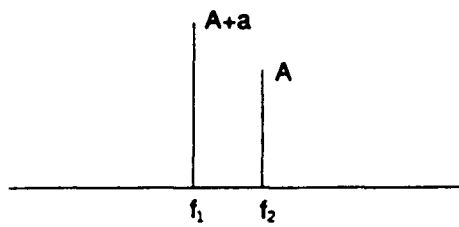
Putting $n = N - m$ gives us

$$e_2(t) = A \left| \sum_{n=0}^N \alpha^n \exp -j n \omega_0 t \right| \quad (17)$$

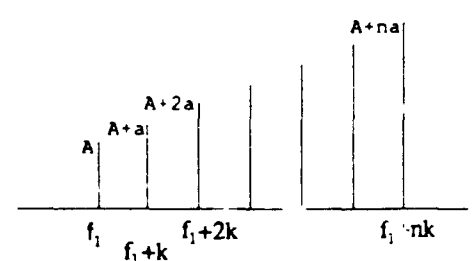
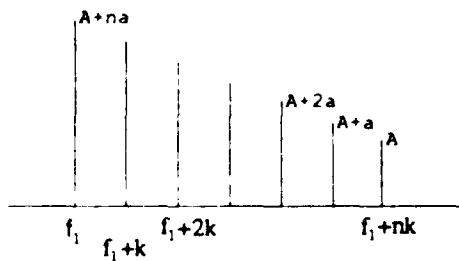
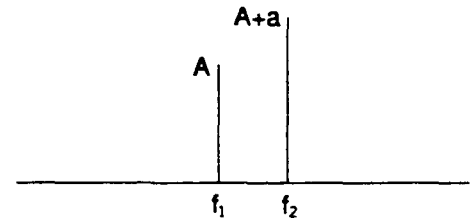
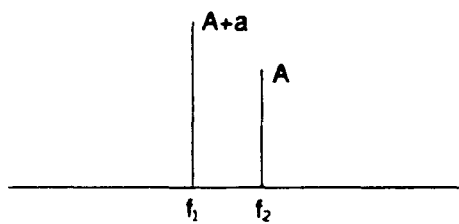
It is easily seen that $e_1(t) = e_2(t)$. Hence the pair described in (10)–(11) specifies another pair of signals having the same envelope functions.

Common envelope signals

Type I :



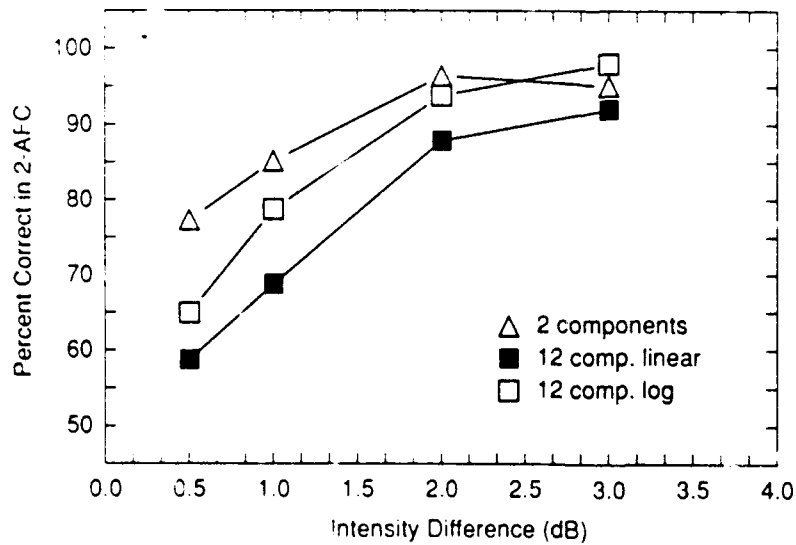
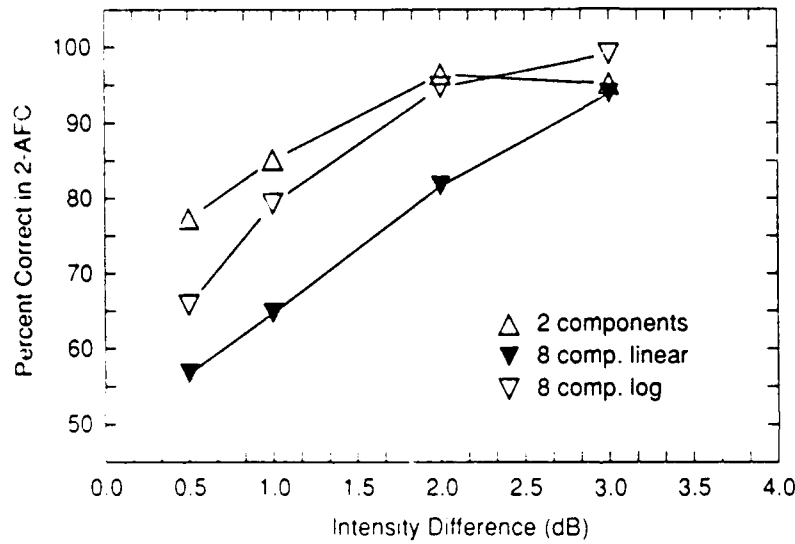
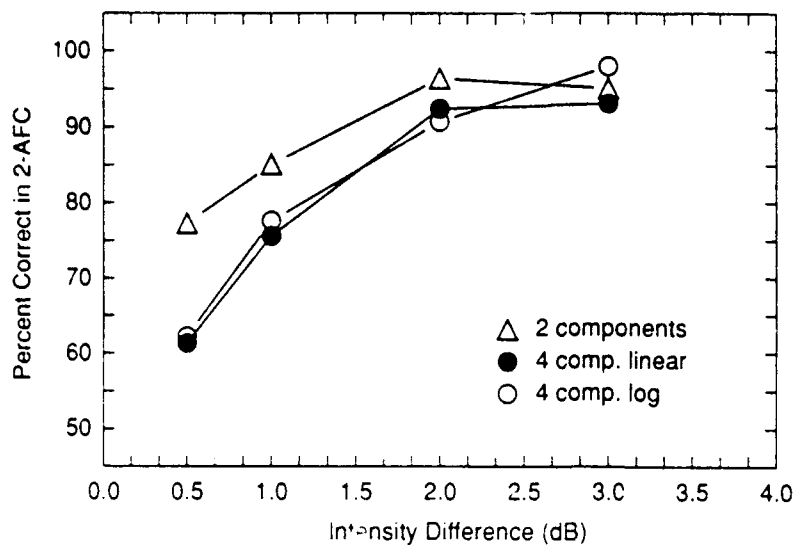
Type II :



Amplitude(dB – log scale)

Frequency (Hz – linear scale)

Fig. 1



COPY AVAILABLE TO DTIC BUT NOT PERMIT FULLY LEGAL
 Approach from Above
 (100msec at 1000Hz)

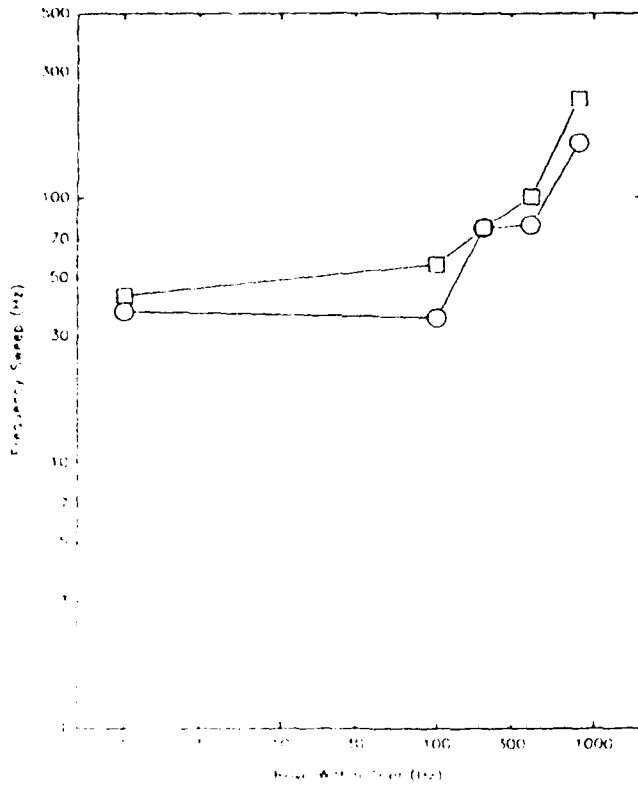
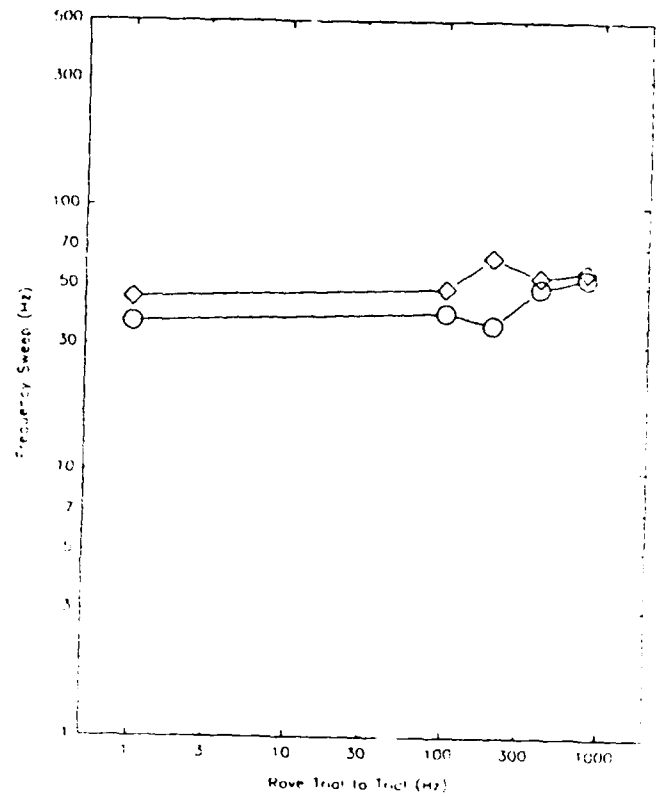
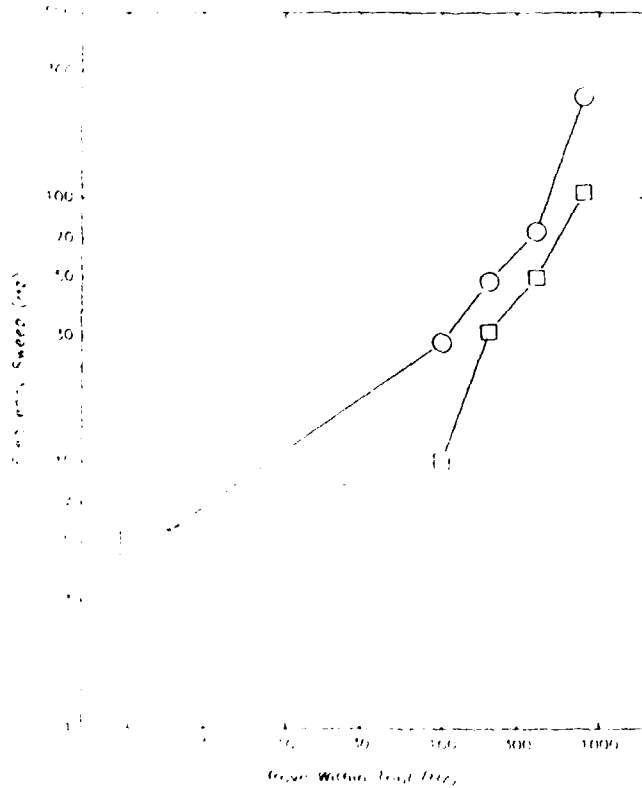


Fig 2

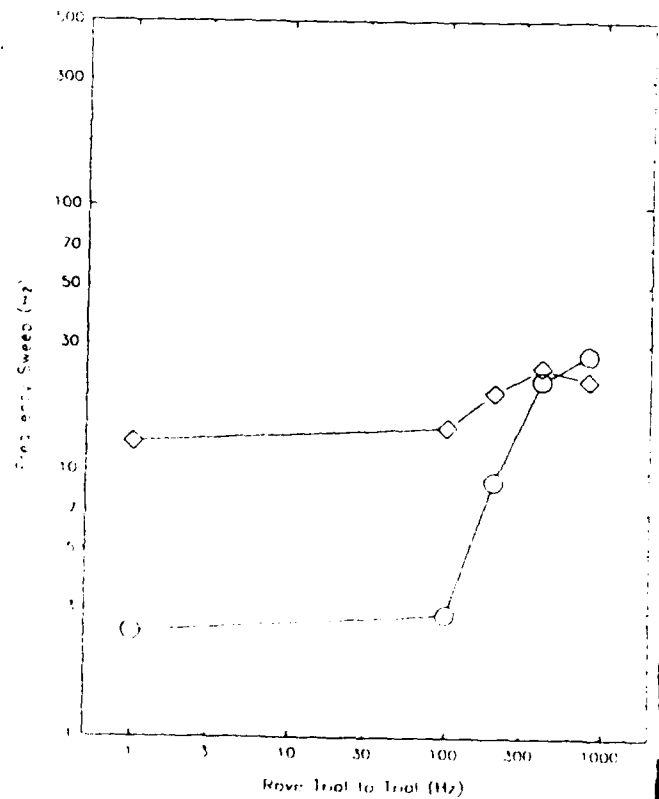
Approach from Above
 (100msec at 1000Hz)



Approach from Below
 (100msec at 1000Hz)



Approach from Below
 (100msec at 1000Hz)



COPY AVAILABLE TO DTIC DOES NOT PERMIT FULLY LEGIBLE REPRODUCTION

Appendices

Temporal Resolution in Normal-Hearing and Hearing-Impaired Listeners Using Frequency-Modulated Stimuli

John P. Madden*
Lawrence L. Feth

*Division of Speech and Hearing Science
The Ohio State University
Columbus, OH*

This study compares the temporal resolution of frequency-modulated sinusoids by normal-hearing and hearing-impaired subjects in a discrimination task. One signal increased linearly by 200 Hz in 50 msec. The other was identical except that its trajectory followed a series of discrete steps. Center frequencies were 500, 1000, 2000, and 4000 Hz. As the number of steps was increased, the duration of the individual steps decreased, and the subjects' discrimination performance monotonically decreased to chance. It was hypothesized that the listeners could not temporally resolve the trajectory of the step signals at short step durations. At equal sensation levels, and at equal sound pressure levels, temporal resolution was significantly reduced for the impaired subjects. The difference between groups was smaller in the equal sound pressure level condition. Performance was much poorer at 4000 Hz than at the other test frequencies in all conditions because of poorer frequency discrimination at that frequency.

KEY WORDS: temporal resolution, frequency modulation, sensorineural hearing loss

Studies of auditory temporal resolution measure the ability of the auditory system to resolve temporal changes in an acoustic signal. A number of approaches have been employed over the last several decades to investigate the temporal resolution of the normal auditory system, among them temporal modulation transfer studies (Viemeister, 1979), forward-masking studies (Nelson & Freyman, 1987; Plomp, 1964), and gap detection studies (Fitzgibbons, 1983; Shailer & Moore, 1985, 1987). For the most part, the experimental task of choice has been gap detection, probably because of its relative convenience. Several issues have been raised in this research that have not been completely resolved. One issue is the effect of hearing impairment. Some studies have found significantly poorer gap-detection thresholds in hearing-impaired listeners in comparison with normal-hearing listeners (e.g., Fitzgibbons & Wightman, 1982), but others have not (Florentine & Buus, 1984). Another issue is the effect of frequency on gap-detection performance. In gap-detection studies using narrow-band noise stimuli, a decrease in gap-detection threshold with increase in frequency has been observed in both normal and impaired ears (Fitzgibbons & Gordon-Salant, 1987). Fitzgibbons argues that the faster response of the more broadly-tuned high-frequency auditory filters accounts for the increase in temporal acuity at the higher frequencies. However, in a study by Moore, Glasberg, Donaldson, McPherson, and Plack (1989) using sinusoidal stimuli, gap-detection thresholds did not change significantly between 400 and 2000 Hz, and the results of Formby and

*Currently affiliated with the Speech and Hearing Department, Cleveland State University, Cleveland, Ohio.

Forest (1991) indicate that gap-detection thresholds are independent of frequency between 500 and 4000 Hz.

Gap-detection tasks, as well as forward-masking and temporal modulation transfer studies, involve the resolution of amplitude changes in experimental stimuli. The literature on temporal resolution is dominated by such studies. There are a few exceptions to this generalization. For example, Jesteadt, Bilger, Green, and Patterson (1976) compared the temporal resolution of the normal and impaired ears of listeners with unilateral hearing losses using Huffman sequences as stimuli (Huffman, 1962). The results indicated that temporal acuity was better for the ear showing the poorer hearing in 8 out of the 10 subjects tested. Studies of the detection of frequency modulation (FM) in sinusoids have implications for temporal acuity (Kay, 1982, presents a review of this research). However, direct investigations of temporal resolution using FM stimuli are relatively rare, despite the ubiquity of FM in naturally occurring sounds such as speech. Studies of temporal acuity in hearing-impaired listeners using FM stimuli are nearly nonexistent.

Feth, Neill, and Krishnamurthy (1989) recently investigated normal temporal resolution with a new discrimination task in which FM stimuli were used. Subjects were asked to discriminate between two sinusoidal signals. One signal, the glide, made a transition from a lower frequency to a higher frequency over a smooth, linear path. The other signal, called the step signal, began and ended at the same frequencies as the glide, but its trajectory followed a series of discrete steps. That is, the signal remained at one frequency for a brief time before abruptly jumping to the next frequency. Normal-hearing listeners were able to distinguish step from glide signals easily when the number of steps was small, but as the number of steps increased, discrimination performance monotonically decreased to chance. It was assumed that at this point the limits of the listener's ability to resolve the discontinuous trajectory of the step signal had been reached, and it was indistinguishable from the glide signal. From the performance of the subject on this task it was possible, therefore, to make inferences about the listener's temporal resolution capacity.

Results obtained by Feth et al. (1989) indicated a temporal resolution threshold of about 6 to 10 msec for normal-hearing listeners at center frequencies from 250 to 2000 Hz, a range that is comparable to estimates of temporal resolution found in gap-detection studies (e.g., Fitzgibbons & Wightman, 1982; Glasberg, Moore & Bacon, 1987). However, resolution at 4000 Hz was much poorer, in the 15–20-msec range. Frequency transitions ranged from 100 to 400 Hz, and signal durations from 25 to 100 msec were used.

The major purpose of the present study was to compare temporal resolution in listeners with moderate hearing losses of presumed sensorineural origin with that of normal-hearing listeners, using FM signals. A second goal was to investigate the effect of frequency on the resolution of FM signals.

Method

Subjects

Five hearing-impaired and 5 normal-hearing listeners participated in the study. The normal-hearing subjects ranged in

age from 20 to 22 years and had pure-tone air-conduction thresholds of less than 15 dB HL (ANSI, 1969) between 500 and 4000 Hz.

The ages and hearing thresholds of the test ears of the hearing-impaired subjects are given in Table 1. All suffered from bilateral sensorineural hearing losses. Hearing thresholds in the nontest ear were no more than 10 dB lower than the test ear thresholds at the respective test frequencies. Bone-conduction testing indicated air-bone gaps of 5 dB or less in all subjects. All hearing losses were long-standing, and there were no indications of retrocochlear involvement in any of the subjects. The hearing losses of H1 and H2 were apparently congenital. H4 reported that her loss was associated with a high fever suffered early in childhood. H3 and H5 indicated that the onset of their hearing losses occurred in adulthood.

Stimuli

Glide and step signals. The glide and step signals were generated and stored in digital form on a Zenith Z159 microcomputer. A 16-bit digital-to-analog converter (Quikki) operating at a 20-kHz sampling rate converted the stored signals to analog waveforms. The resulting signals were low-pass filtered at 8000 Hz. The glide signals were sinusoidal sweep tones with center frequencies of 500, 1000, 2000, and 4000 Hz. The frequency transition was 200 Hz over 50 ms, producing a 4-Hz/ms rate of frequency change. Rise/fall time was 5 msec, resulting in an overall signal duration of 60 msec. Signal onsets and offsets were shaped by a cosine-squared function. The signal duration was chosen to approximate the duration of formant transitions in the speech signal. The step signals traversed the same frequency range as the glide signals, but did so in discrete steps. The number of steps varied between two and nine. Schematic representations of the glide and step signals are shown in Figure 1.

Spectral analysis. Abrupt frequency jumps such as those in the step signal generate off-frequency spectral energy, which is a potential discrimination cue. To minimize this potential confounding variable, the step signals were generated with rounded "corners." Spectral analysis indicated that the long-term spectra of the step signals were essentially identical to that of the glide signal. Figure 2 shows a comparison of the long-term electrical spectra of a four-step signal and a glide signal. The acoustical spectra of the signals were essentially identical to their electrical spectra.

Procedures

Signal levels. The hearing-impaired subjects were tested

TABLE 1. Hearing-impaired subject information. Hearing thresholds are in dB HL.

Subject	Age	500 Hz	1000 Hz	2000 Hz	4000 Hz
H1	24	30	35	40	40
H2	25	35	35	50	55
H3	74	45	45	45	65
H4	27	35	50	60	60
H5	44	50	60	60	55

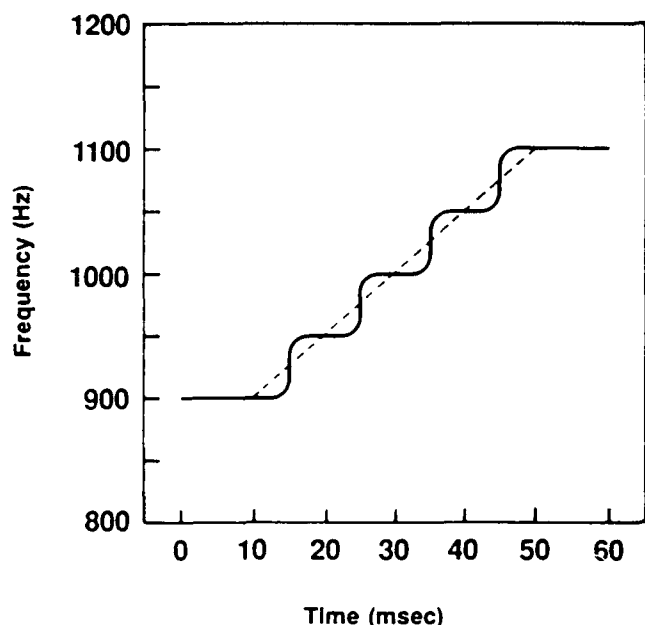


FIGURE 1. Schematic representation of the step and glide signal trajectories. The solid line represents the step signal; the broken line represents the glide signal.

in quiet at a sensation level (SL) of 35 dB. The normal-hearing subjects were tested in two conditions. In the first condition, the quiet condition, the signals were presented in quiet at 35 dB SL. Feth et al. (1989) found that discrimination performance in normal-hearing subjects is asymptotic at intensities as low as 30 dB SL. In the second condition, the masked condition, the stimuli were presented at sound pressure levels (SPLs) that approximated the average levels used for the impaired ears and were as follows: 500 Hz: 75 dB; 1000 Hz: 80 dB; 2000 Hz: 82 dB; and 4000 Hz: 83 dB. SPLs were determined using a flat-plate coupler. Broadband masking noise was low-pass filtered at 8000 Hz and combined with the signal to achieve a signal SL of approximately 35 dB. Thus, in the two conditions the normal-hearing subjects were compared to the hearing-impaired subjects at

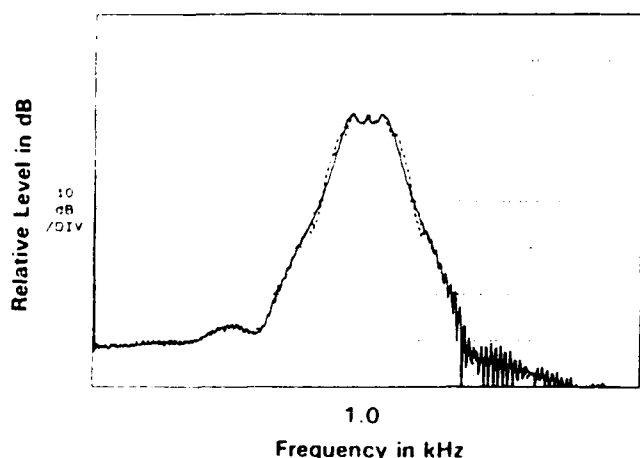


FIGURE 2. Long-term frequency spectrum of the glide signal (solid line) and a four-step signal (broken line).

both an equal SL (the quiet condition) and at an equal SPL and equal SL (the masked condition).

Data collection. Subjects were tested in a single-walled sound-attenuating chamber. Stimuli were presented monaurally via Sennheiser HD414SL headphones. A two-cue, two-alternative forced-choice procedure (2-Q, 2AFC) was used to determine step/glide discrimination performance. In each trial a stimulus was presented in each of four listening intervals. The interstimulus interval was 400 msec. The subject was asked to pick the odd stimulus, which was always the step signal and was always presented, randomly, in interval two or three. Feedback indicating the interval containing the step signal was provided.

The stimuli were presented in blocks of 50 trials, and a percent-correct score was calculated for each block. Each subject was tested over at least three sets of three blocks for each different step signal (nine blocks altogether, or 450 trials). For example, at least nine blocks were run for the 1000-Hz two-step signal, at least nine blocks for the 1000-Hz three-step signal, etc. If the subject showed no improvement in percent correct over the last two sets of blocks, data collection was ended for that signal. The percent-correct score for that step duration (step duration being a function of the number of steps in the signal) was obtained by calculating the mean of the percent-correct scores from each of the last six blocks. In the few cases where improvement continued, additional sets of three blocks were run until no further improvement was found. Percent correct was calculated as described above for the last six blocks. Thus, each data point in the individual results is based on 300 discrimination trials. The data points were used to construct psychometric functions for each of the test frequencies in each of the experimental conditions.

All of the subjects readily learned the procedure except for the oldest subject, H3, who had difficulty with the rate of stimulus presentation. However, when stimulus presentation was slowed to one half its normal rate, the subject quickly mastered the task. All subjects were well practiced in the task when data collection was begun.

Results

The psychometric functions in Figure 3 display the mean discrimination results at center frequencies from 500 to 4000 Hz for all conditions. Percent-correct discrimination is plotted as a function of step duration of the step signal. The open symbols indicate data obtained from the normal-hearing and the hearing-impaired subjects in the quiet condition. The filled symbols represent data obtained from the normal-hearing subjects in the masked condition. The temporal resolution threshold (TRT) was defined as 75% correct discrimination. The mean TRT for each condition at the various test frequencies is given in Table 2. The threshold values are the points on the x-axis at which the psychometric functions intercept the 75% correct level, estimated to the nearest 0.5 ms.

Table 2 indicates that the mean TRTs of the hearing-impaired listeners are poorer than those of the normal-hearing listeners at all frequencies. No comparison is possi-

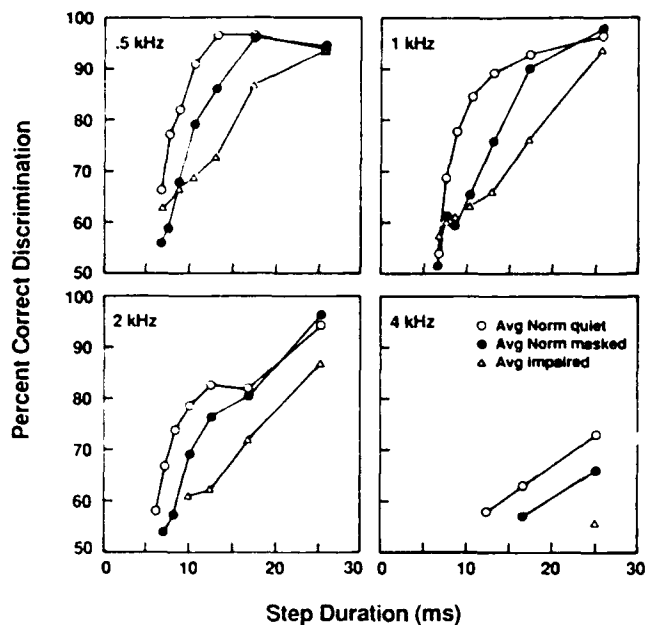


FIGURE 3. Psychometric functions showing the mean discrimination results. The open circles represent results obtained from the normal subjects at a signal level of 35 dB SL. The filled circles represent the results from the normal subjects at a signal level approximating that of the impaired subjects with broadband masking noise added to produce a sensation level of 35 dB. The triangles are the data from the impaired subjects at a signal level of 35 dB SL. The horizontal lines mark the 75% correct discrimination point.

ble at 4000 Hz because the maximum mean discrimination scores were below the 75% criterion for both groups. It is clear that the differences between the two groups are substantial, even in the masked comparison. It is also evident from Table 2 that temporal resolution is poorer at every frequency in the masked condition than it is in the quiet condition for the normal-hearing subjects.

Several analysis of variance tests were performed to support the conclusions reached through visual inspection of the data. The difference in TRTs between the normal and impaired listeners, with group as a between-subjects factor, was significant when the normal-hearing subjects were compared to the impaired subjects in both the quiet [$F(1,8) = 13.90, p < .006$] and the masked [$F(1,8) = 5.52, p < .047$] conditions. Also, a comparison of the results from the normal-

TABLE 2. Mean temporal resolution thresholds for each of the experimental conditions.

Condition	500 Hz		1000 Hz		2000 Hz		4000 Hz	
	M	SD	M	SD	M	SD	M	SD
Normal quiet	7.0	1.2	8.0	2.3	9.0	3.5	25.0*	
Normal masked	9.5	1.5	12.0	1.0	12.0	3.0		
Impaired	13.0	3.3	16.0	3.8	18.0	3.7		

Note. Values were obtained by visual inspection of Figure 3 and are estimated to the nearest 0.5 msec. A mean discrimination score of 75% correct was not achieved or approximated at the longest step duration in the normal masked and impaired conditions at 4000 Hz. *Maximum mean discrimination was actually 73%.

hearing subjects in the masked and the quiet conditions, with masking treated as a within-subjects factor, indicated that the effect of condition was significant [$F(1,4) = 37.24, p < .004$].

Table 2 also indicates that there is a dramatic increase in TRT at 4000 Hz in all three conditions. Even at the longest step duration (25 msec), only 3 of the normal subjects were able to achieve the 75% correct discrimination criterion at 4000 Hz in the quiet condition. None of the normal subjects reached the criterion value at 4000 Hz in the masked condition. The impaired subjects also failed to achieve 75% correct discrimination at 4000 Hz. In the case of the normal-hearing listeners, TRTs increased very slightly between 500 and 2000 Hz in the quiet and masked conditions. However, there is a considerably greater increase in threshold between 500 and 2000 Hz for the impaired subjects.

One-way analysis of variance tests with frequency as a within-subjects variable were performed on the data from 500, 1000, and 2000 Hz. The 4000-Hz data were not included because too few of the subjects reached the discrimination criterion at that frequency. The results indicated that the effect of frequency was not significant over this frequency range in the case of the normal-hearing subjects in quiet or in noise. However, the effect of frequency over the 500–2000-Hz range was significant in the case of the impaired subjects [$F(2,8) = 4.95, p < .04$]. This effect can be accounted for in terms of hearing sensitivity. In Figure 4, TRT is plotted as a function of hearing threshold level for the hearing-impaired subjects. A strong positive relation between TRT and hearing threshold is evident, and this is confirmed by correlational analysis ($r = 0.68, p < .01$). And, in general, the hearing thresholds of the hearing-impaired subjects increase with frequency. Therefore, it appears that the apparent increase in TRT with frequency is in fact an increase in TRT as hearing threshold increases.

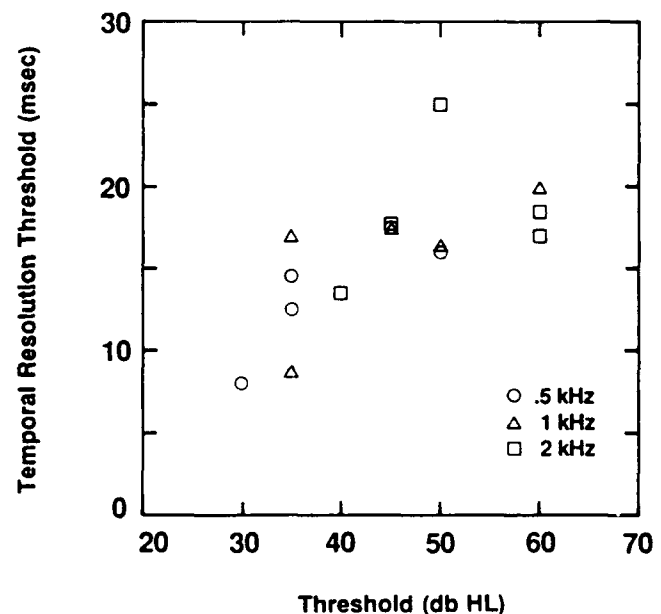


FIGURE 4. Temporal resolution of the individual hearing-impaired subjects plotted as a function of hearing threshold. The parameter is stimulus center frequency.

In summary, the effect of frequency on TRT appears limited to 4000 Hz. There is no statistically significant effect of frequency between 500 and 2000 Hz in the case of the normal-hearing subjects. The increase in TRT between 500 and 2000 Hz observed in the hearing-impaired subjects appears to be a function of increasing hearing threshold in the higher frequencies. In marked contrast to the results in the lower frequencies, discrimination performance in all conditions was so poor at 4000 Hz that the mean discrimination scores failed to reach the criterion value even at the 25-msec step size.

Overall, the individual results follow the trends described above for the averaged data, with one exception. Normal subject N5's TRTs are poorer than those of any of the other normal-hearing subjects at nearly all test frequencies in both the quiet and masked conditions. For example, N5's TRTs in the quiet condition are as follows: 500 Hz: 9.5 msec; 1000 Hz: 13.0 msec; 2000 Hz: 16.5 msec. N5's results were consistent, and the subject performed well in other aspects of the experiment, such as the hearing threshold measurements, indicating that lack of concentration or motivation is not a likely explanation for these results. In Feth et al.'s (1989) results from normal subjects, no listener's thresholds departed this far from the mean. The data from N5 suggest that some individuals with normal hearing sensitivity may have abnormally large temporal resolution thresholds.

Discussion

The Effect of Frequency

In the case of the normal listeners in quiet, there is no significant variation in mean temporal resolution threshold between 500 and 2000 Hz, but at 4000 Hz the TRT increases to greater than 25 msec. This pattern is not seen in studies using other measures of temporal acuity. Formby and Forrest (1991), for example, found that gap detection thresholds measured with sinusoidal markers are independent of frequency from 500 to 4000 Hz. One possible explanation for the increase in TRT at 4000 Hz is that the auditory system tracks the frequency changes in the step signal using information from phase-locked neural discharges. If this were true, then temporal resolution would be expected to deteriorate as phase-locking declines. It is well known that in monkeys and cats, neural phase-locking is robust below 1 kHz, declines gradually at higher frequencies, and is absent above 4000 to 5000 Hz (Rose, Brugge & Hind, 1967). The frequency effect observed in the normal subjects fits this pattern. One might infer that the mechanism that takes over signal-tracking (perhaps rate-place coding) has a longer time constant than the phase-locking mechanism.

There is another less speculative explanation, however. The step signal is, in effect, a sequence of level tones separated by almost instantaneous frequency transitions. As the step duration decreases, the extent of the frequency transition between the steps decreases as well. It can be argued, therefore, that the subject's frequency discrimination ability is the limiting factor in the task, rather than the subject's temporal acuity. Feth et al. (1989) investigated the

role of frequency discrimination in the step-glide discrimination task with normal-hearing subjects. They varied the extent of frequency transition, using signals with transitions of 200 and 400 Hz while holding the length of the signal constant at 50 msec. If frequency discrimination is the limiting factor in the discrimination task, then the subjects' performance should improve for the 400-Hz transition signals, in which the between-step jumps are twice those of the 200-Hz transition signals. At center frequencies of 500, 1000, and 2000 Hz, there was no significant improvement in mean temporal resolution threshold when the transition size was increased. The TRTs of the 200-Hz signals were within 0.5 ms of the TRTs of the 400-Hz signals. These data support the contention that frequency discrimination does not play a limiting role for signals with frequency transitions of 200 Hz at center frequencies of 2000 Hz and below.

However, Feth et al. found that at a center frequency of 4000 Hz, the TRT obtained for the 200-Hz transition signal was about 7 msec greater than the TRT for the 400-Hz transition signal. These data strongly suggest that frequency discrimination has a considerable effect on performance at 4000 Hz, where the frequency DL is relatively large (Moore, 1973; Wier, Jesteadt, & Green, 1977). Thus, the poor step-glide discrimination observed at 4000 Hz in the present study probably reflects the effect of frequency discrimination rather than temporal resolution. In the Feth et al. study, a step signal without rounded corners was used, and TRTs were smaller, particularly at 4000 Hz. Nevertheless, the TRTs obtained at the lower frequencies are very similar to those of the present study, and the two studies are highly similar with respect to the overall pattern of their results.

The Effect of Frequency Discrimination in the Hearing-Impaired Subjects

In the normal-hearing listeners, frequency discrimination appears to affect step-glide discrimination only at 4000 Hz. However, it may be argued that the poorer temporal resolution of the hearing-impaired subjects in comparison with the normal-hearing listeners also is due to poorer frequency discrimination. To investigate this possibility, difference limens for frequency (DLFs) were obtained for 3 of the normal-hearing and 3 of the hearing-impaired subjects, and the correlation between DLFs and TRTs for these subjects was obtained. A strong relationship between these two variables would indicate that frequency discrimination is a major determining factor in the poorer performance of the hearing-impaired subjects.

To measure DLFs, 50-msec sinusoids (5 msec rise/fall time) were presented at 35 dB SL in the same 2-Q, 2AFC task that was used for the step/glide discrimination task. An adaptive procedure was used that estimated the 70.7% correct point on the psychometric function (Levitt, 1971). Table 3 displays the DLFs of the subjects tested. Correlational analysis indicated that there is no relation between the DLFs and TRTs of these subjects ($r = .017$). It therefore seems unlikely that frequency discrimination played a major role in limiting the performance of the hearing-impaired subjects.

TABLE 3. DLFs in Hz for normal-hearing and hearing-impaired listeners.

Subject	500 Hz	1000 Hz	2000 Hz
<i>Normal-hearing</i>			
N2	3.0	5.1	31.6
N3	7.5	15.5	25.6
N4	4.8	7.1	9.2
Avg	5.1	9.2	22.0
<i>Hearing-impaired</i>			
H2	6.6	10.0	7.2
H4	3.1	6.8	14.5
H5	5.4	16.6	29.3
Avg	5.0	11.1	17.0

The frequency discrimination results are similar to those obtained in other studies. Hall and Wood (1984), also using 50-msec signals, obtained the following results: at 500 Hz, normal-hearing subjects: 1.2–4.2 Hz; impaired subjects, 4.1–21.0 Hz. At 2000 Hz, normal-hearing subjects: 1.9–9.6 Hz; impaired subjects: 4.3–25.7 Hz. The somewhat higher values in the present results are probably due to differences in presentation level (Hall & Wood used 90 dB SPL) and practice time. Hall and Wood's subjects were given 4 hours of practice, whereas these subjects received less than 2 hours. The intent here was not to obtain optimal results, but to obtain comparable results. The fact that the hearing-impaired subject results are very similar to the results from the normal-hearing subjects is somewhat surprising but not unprecedented. Tyler, Wood, and Fernandes (1983), commenting on their discrimination results, remark that many subjects with hearing thresholds greater than 60 to 70 dB SPL display normal frequency discrimination.

The Effect of Stimulus Level

The performance of the normal-hearing subjects is poorer in the masked condition than in the quiet condition. This result prompts one to ask whether the degradation of performance in the masked condition is due to the presence of masking noise or to the higher level at which the stimuli are presented in the masked condition. Preliminary step-glide discrimination data, obtained in 3 of the normal-hearing subjects by presenting the stimuli at the higher SPLs without the addition of masking noise, suggest that level is the controlling variable. The discrimination performance of all 3 subjects was found to decrease as stimulus level was increased. This decline in performance with increasing level is unusual in psychoacoustic phenomena. In general, performance increases in auditory discrimination tasks until at least 80 dB SPL. This is the case, for example, for gap detection (Plomp, 1964) and frequency discrimination (Wier et al., 1977).

The question that now arises is whether the effect of absolute level is the same for both the impaired and the normal subjects. That is, to what extent can the poorer temporal resolution of the impaired subjects be accounted for by their higher listening level? Obviously, a difference in stimulus level cannot account entirely for the difference between the two groups. In the

masked condition, the normal subjects are compared to the hearing-impaired group at approximately equal SPLs, and the performance of the impaired subjects is poorer. Two of the hearing-impaired subjects, H2 and H4, were tested at a range of levels. These subjects' discrimination performance begins to decline at about 20 to 25 dB SL as stimulus level is decreased, and at about 35 dB SL as stimulus level is increased. That is, optimal temporal resolution was obtained at 25 to 35 dB SL in the hearing-impaired subjects tested. (This indicates that an optimal listening level was at least approximated for the impaired listeners in the study.) One interpretation of these data is that stimulus audibility imposes the lower cut-off point for optimal performance and that the higher cut-off point is determined by an intensity-discrimination function that is similar to that of the normal ear. A problem with this explanation is the fact that signals at sensation levels less than 25 dB ought to be quite audible to the hearing-impaired subjects because of the rapid growth of loudness at higher levels typically found in individuals with hearing impairment of sensorineural origin (Sanders, 1979). Fitzgibbons (1984) argues that a stimulus of 20 dB SL should be more than sufficiently loud to elicit optimal gap detection results in hearing-impaired ears, despite the decrease in performance below 30 dB SL in normal-hearing ears. In any event, it appears that the dynamic range of the hearing-impaired subjects is severely limited for the experimental task in this study. Further research is needed to establish the exact nature of this limitation.

Conclusions

The major findings of the study, as observed in normal-hearing subjects and subjects with mild-to-moderate sensorineural hearing losses, are as follows:

1. When normal-hearing and hearing-impaired subjects were compared at equal sensation levels, mean temporal resolution thresholds were significantly greater in the hearing-impaired subjects. There was a strong positive correlation between temporal resolution threshold and hearing threshold for the hearing-impaired listeners.
2. When the normal-hearing and hearing-impaired subjects were compared at equal sensation levels and equal sound pressure levels, the mean temporal resolution threshold of the hearing-impaired subjects was also significantly greater than that of the normal-hearing subjects, but the difference was smaller than it was in the equal sensation level condition.
3. Temporal resolution thresholds were essentially independent of frequency at 500, 1000, and 2000 Hz. Step/glide discrimination was much poorer at 4000 Hz than at the other test frequencies, apparently because of poorer frequency discrimination at that frequency.

Acknowledgments

The authors wish to thank Dr. Ying-Yong Qi and Chien-Yeh Hsu for their programming assistance, and Mary Neill for her help in data collection. They also wish to express their appreciation to Peter J. Fitzgibbons, C. Formby, and Richard H. Wilson for their helpful comments.

This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR-89-0227. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon.

References

- American National Standards Institute.** (1969). *Specifications for audiometers* (ANSI S3.6-1969). New York: ANSI.
- Feth, L., Nelli, M. E., & Krishnamurthy, A.** (1989). Auditory temporal acuity for dynamic signals. *Journal of the Acoustical Society of America*, 86, S122-S123.
- Fitzgibbons, P. J.** (1983). Temporal gap detection in noise as a function of frequency, bandwidth, and level. *Journal of the Acoustical Society of America*, 74, 67-72.
- Fitzgibbons, P. J.** (1984). Temporal gap resolution in masked normal ears as a function of masker level. *Journal of the Acoustical Society of America*, 76, 67-70.
- Fitzgibbons, P. J., & Gordon-Salant, S.** (1987). Temporal gap resolution in listeners with high-frequency sensorineural hearing loss. *Journal of the Acoustical Society of America*, 81, 133-137.
- Fitzgibbons, P. J., & Wightman, F. L.** (1982). Gap detection in normal and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 72, 761-765.
- Florentine, M., & Buus, S.** (1984). Temporal gap detection in sensorineural and simulated hearing impairments. *Journal of the Acoustical Society of America*, 77, 449-455.
- Formby, C., & Forest, T. G.** (1991). Detection of silent temporal gaps in sinusoidal markers. *Journal of the Acoustical Society of America*, 89, 830-837.
- Glasberg, B. R., Moore, B. C. J., & Bacon, S. P.** (1987). Gap detection and masking in hearing-impaired and normal-hearing subjects. *Journal of the Acoustical Society of America*, 81, 1546-1556.
- Hall, J. H., & Wood, E. J.** (1984). Stimulus duration and frequency discrimination for normal-hearing and hearing-impaired subjects. *Journal of the Acoustical Society of America*, 77, 252-256.
- Huffman, D. A.** (1962). The generation of impulse-equivalent pulse trains. *IRE Transactions*, IT8, 510-516.
- Jesteadt, W., Bilger, R. C., Green, D. M., & Patterson, J.** (1976). Temporal acuity in listeners with sensorineural hearing loss. *Journal of Speech and Hearing Research*, 19, 357-370.
- Kay, R. H.** (1982). Hearing of modulation in sounds. *Physiological Reviews*, 62, 894-975.
- Levitt, H.** (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467-477.
- Moore, B. C. J.** (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, 54, 610-619.
- Moore, B. C. J., Glasberg, B. R., Donaldson, E., McPherson, T., & Plack, C.** (1989). Detection of temporal gaps in sinusoids by normally hearing and hearing-impaired subjects. *Journal of the Acoustical Society of America*, 85, 1266-1275.
- Nelson, D. A., & Freyman, R. L.** (1987). Temporal resolution in sensorineural hearing-impaired listeners. *Journal of the Acoustical Society of America*, 81, 709-720.
- Plomp, R.** (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of America*, 36, 277-282.
- Rose, J. E., Brugge, J. F., & Hind, J. E.** (1967). Phase locked response to low frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of Neurophysiology*, 30, 769-793.
- Sanders, J. W.** (1979). Recruitment. In W. F. Rintelmann (Ed.), *Hearing assessment* (pp. 261-280). Baltimore: University Park Press.
- Shaller, M. J., & Moore, B. C. J.** (1985). Detection of temporal gaps in bandlimited noise: Effects of variations in bandwidth and signal-to-noise ratio. *Journal of the Acoustical Society of America*, 77, 635-639.
- Shaller, M. J., & Moore, B. C. J.** (1987). Gap detection and the auditory filter: Phase effects using sinusoidal stimuli. *Journal of the Acoustical Society of America*, 81, 1110-1117.
- Tyler, R. S., Wood, E. J., & Fernandes, M.** (1983). Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 74, 1190-1199.
- Vlamester, N. F.** (1979). Temporal modulation transfer functions based upon modulation thresholds. *Journal of the Acoustical Society of America*, 66, 1354-1380.
- Wier, C. C., Jesteadt, W., & Green, D. M.** (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, 61, 178-184.

Received February 1, 1991

Accepted June 7, 1991

Contact author: John P. Madden, PhD, Speech and Hearing Department, Cleveland State University, Euclid Avenue at East 24th Street, Cleveland, OH 44115.

Auditory Temporal Acuity Using Frequency Modulated Sinusoids

Lawrence L. Feth and John P. Madden¹

Speech and Hearing Science

Ohio State University

Columbus, OH 43210

Submitted to the

Journal of the Acoustical Society of America

April 1992

¹Present Address: Speech and Hearing Department

Cleveland State University

Cleveland, OH 44115

Temporal Acuity with FM Glides

Abstract

A means of determining the temporal acuity of the human auditory system using frequency-modulated (FM) signals is proposed. These FM signals have some characteristics in common with the formant transitions of speech, and thus may be useful in relating psychoacoustic performance to speech processing. Listeners with normal hearing were asked to discriminate between a sinusoid modulated linearly over a brief time interval (a GLIDE), from a signal covering the same frequency excursion in a multiple-STEP trajectory. When the GLIDE signal and the STEP signal were just discriminable (75% correct in 2Q,2AFC) we assumed that the duration of a single step was less than the width of the auditory temporal window. For presentation frequencies from 250 Hz through 2000 Hz, the estimate of the width of the temporal window was 7 to 10 msec. Presentation level had no effect on results, at least for 30 and 50 dB SL. Testing above 50 dB SL was limited because spectral differences between test signals might confound the results. Above 2 kHz, performance was poorer in the discrimination task. In fact, at 4 kHz we cannot rule out the possibility that listeners were basing their decisions on just discriminable frequency jumps, rather than on temporal differences. A follow up at 6 kHz using different listeners led to equivocal results.

Introduction

Studies of auditory temporal resolution measure the ability of the auditory system to follow rapid changes in an acoustic signal. Estimates of temporal resolution have been derived from several different experimental tasks which use signals that change rapidly in amplitude. Examples include forward masking, temporal modulation transfer function (TMTF), and gap detection studies. The results of these studies have been relatively consistent: gap thresholds obtained with broadband noise markers are on the order of 3 to 5 msec (e.g., Plomp, 1964; Florentine and Buus, 1984), and minimum temporal integration times observed in TMTF studies are in the 2 to 8 msec range (Scott, 1986). In both gap detection and TMTF studies, temporal acuity appears to be relatively independent of level over a wide range, declining significantly only at low levels, evidently as audibility decreases (Buus and Florentine, 1985; Viemeister, 1979).

Gap thresholds for narrowband noise markers decrease as the center frequency of the stimulus increases. A similar effect has been reported for TMTF studies (Fitzgibbons, 1979; Fitzgibbons and Wightman, 1982; Scott, 1986). Initially, it was hypothesized that this effect was due to the shorter response time of the more broadly tuned auditory filters in the higher frequencies. However, there is now considerable evidence that this effect is due to fluctuations in level that are inherent in narrowband noise (Moore and Glasberg, 1977; Shailer and Moore, 1987; Moore, 1989; Eddins, et al., 1992). In the earlier gap detection experiments, octave-band noise was used, and thus the bandwidth of the stimuli decreased with decreasing center frequency. The slower fluctuations in level at the narrower bandwidths were more easily confused with the gaps, causing greater gap thresholds at the lower frequencies. In fact, when deterministic signals, which contain no level

fluctuations, are used as markers, gap thresholds are essentially independent of frequency (Shailer and Moore, 1987; Formby and Forrest, 1991).

Several investigators have proposed a model to account for gap detection and TMTF results. The model requires a bank of auditory filters. The filter bank is followed by a non-linear device, a temporal window, and a level detector acting on the output of each channel (Shailer and Moore, 1987; Green and Forest, 1988). Because the temporal window seems to play the major role in determining temporal resolution characteristics of the auditory system, there have been a series of attempts to provide a simple estimate of its width. Drawing on the analogy with the critical band, these studies initially attempted to determine the width of a critical masking interval (Penner, Robinson, and Green, 1972; Penner and Cudahy, 1973; Robinson and Pollack, 1973, Penner et al., 1974). The results of these studies were inconsistent, at least in part because of a failure to control for the effects of "off-time" listening (Moore, Glasberg, Plack and Biswas, 1988). Moore and his colleagues (Moore et al., 1988; Plack and Moore, 1990) appear to have avoided this problem in experiments that measure the threshold of a brief sinusoid presented in a temporal gap between two noise bursts. Their results led them to describe the temporal window as an asymmetric temporal intensity-weighting function that could be modeled well by fitting rounded exponentials to each side of the function. Plack and Moore's (1990) estimate of the equivalent rectangular duration (ERD) of this "roex" temporal window decreased from 13 to 9 msec as center frequency increased from 300 to 900 Hz and then remained relatively constant, declining slightly to 7 msec at 8100 Hz. They also noted an effect of level. Above 900 Hz, level increases produced somewhat narrower ERD estimates. Moore et al. (1988) demonstrate that the calculated output from a window of this general form gives reasonably accurate predictions for at least some aspects of various temporal phenomena involving amplitude modulated signals; for example, the detection

of amplitude changes (Buunen and Valkenburg, 1979) and temporal modulation transfer functions. It should be noted, however, that Shailer and Moore (1987) suggested that a temporal integrator with a 15 msec time constant could account for the absence of marked increases in gap thresholds at low frequencies that might be expected from ringing of the auditory filters. Apparently, there is still some uncertainty as to the exact nature of the temporal window.

In contrast to the multiplicity of studies with steady-state signals, the literature contains relatively few studies concerning the temporal resolution of signals that change rapidly in frequency or phase, despite the ubiquity of such signals in natural sounds such as speech. Patterson and Green (1970) measured temporal acuity using Huffman (1962) sequences, brief waveforms in which energy is delayed in some frequency region. Green (1973) found that subjects could detect a delay time of about 2 msec at 650, 1900, and 4000 Hz. Modulation-rate transfer functions for frequency-modulated (FM) sinusoids can be interpreted as measuring the ability of the auditory system to follow periodic spectral changes. For a carrier frequency of 1 kHz, the detectability of FM sinusoids monotonically decreases at modulation rates greater than about 2 Hz, as is indicated by increased modulation depth needed for detection (Kay, 1982). Detection then improves at rates higher than 100 Hz, apparently due to the resolution of spectral sidebands. At modulation rates of less than 5 to 10 Hz, the frequency changes can be followed perceptually. Between 10 Hz and 100 Hz the sound is described as "rough" or "motorboating" (Kay, 1982). These findings suggest that there is some degree of temporal resolution of sinusoidal FM up to at least 100 Hz.

The purpose of this study was to investigate further the temporal resolution of FM signals using a new experimental task. Subjects were asked to discriminate between two frequency-modulated, sinusoidal signals. One signal, the GLIDE, made a transition from a lower frequency to a higher frequency over a smooth, linear path. The target

Temporal Acuity with FM Glides

signal, called the STEP, was identical to the GLIDE, except that its trajectory followed a series of brief steps. That is, the STEP signal remained at one frequency for a brief time before jumping to the next frequency. Pilot studies indicated, not surprisingly, that as the number of steps was increased, discrimination became more difficult, and listener performance monotonically decreased to chance. It is assumed that discrimination threshold is reached when the listener's percentage of correct discriminations falls below 75%. We infer that the temporal window must be wider than the size of an individual step in the STEP signal at this point. That is, changes in the signal which occur in less time than the width of the temporal window are not distinguishable by the listener.

This task may be a more direct measure of the purely temporal properties of the auditory system than studies that involve the detection of amplitude changes. The current model of temporal resolution implies that gap detection performance, for example, depends on the sensitivity of the level detector, as well as the properties of the temporal window (Moore et al., 1989). Level detector sensitivity may play a major role in the case of hearing-impaired subjects. As long as the frequency jumps between steps are large enough to be easily discriminable, no such confounding factors are present in this paradigm.

The study was designed to investigate the effects of several variables on the temporal resolution of FM signals, including signal duration, signal transition rate, transition size, center frequency, and presentation level.

Methods

Stimuli

The signals used in the discrimination task were frequency-modulated sinusoids. The frequency of the GLIDE changed linearly over a brief time interval, T . The extent of the frequency excursion is labeled ΔF . Thus, GLIDE signals traversed ΔF Hertz in T msec. The STEP signal at each center frequency, f_c , covered the same frequency excursion over the same duration as a GLIDE, but its frequency followed a multiple-step trajectory. The simplest of the STEP signals would remain at the initial frequency for $T/2$ msec, then jump to the final frequency for the remainder of the signal. Other signals would cover ΔF in three, four or more equal-duration steps. Trajectories and long-term spectra for a GLIDE and a 4-STEP signal are shown in Madden and Feth (1992).

Figure 1 presents a display of the response of the auditory filter bank to various STEP and GLIDE signals. Each signal covers 400 Hz in 100 msec at a center frequency of 1 kHz. The filter bank response is taken from the first stage of the Auditory Sensation Processing (ASP) model of Patterson et al. (1992)

Figure 1 about here

GLIDE and STEP signals were generated off-line by a laboratory microcomputer and stored on hard disk for use in each discrimination run. They were converted to analog form at a 20 kHz sampling rate using a 16-bit D-to-A converter (TTES Quikki board). The post D-to-A filter was set to a low pass cutoff frequency of 8.5 kHz. Signals were generated with 5-msec rise and fall times which were shaped by raised cosine functions in the generation program. Frequency transitions did not extend into the rise and fall portions of either signal.

Temporal Acuity with FM Glides

Signals were generated at center frequencies of .25, .5, 1, 2, 4 and 6 kHz. ΔF s covered 100, 200 and 400 Hz at each center frequency. Values for T were 100, 50 and 25 msec, plus 5 msec of rise and fall time.

Subjects

Subjects for this experiment were eight university students who were recruited to serve in the study. All had hearing within normal limits and negative otological histories. They ranged in age from 18 to 26 years. All were female. Listeners were paid an hourly wage for their participation.

Procedures

Testing was conducted in a four-interval, two-alternative procedure, commonly called 2Q,2AFC. GLIDE signals were always presented in the first and fourth intervals. The target signal, STEP, was presented in either the second or the third interval, with equal probability, and a GLIDE signal was presented in the remaining interval. The listeners were instructed to indicate whether interval two or three contained the "odd" signal. They were given feedback to indicate the correct response after each trial. Three blocks of fifty trials each, for a given pair of GLIDE and STEP signals, were run in succession, and each listener's percent correct score for each block was recorded. New signals were selected for the next three-block set. Listeners were given brief rests after each three block run, and longer breaks about every half hour. Testing was usually conducted for three listeners at one time for a period of two hours per day. Results were based on at least six fifty-block trials with no more than 150 trials for a given signal pair collected in one day.

Detection thresholds were determined for each of the GLIDE signals used as standards in the discrimination testing. Listeners were tested in a simple, adaptive 2AFC detection task. Once thresholds were determined at each f_c , discrimination testing was conducted with signals presented at 50 dB SL for each individual listener.

Results

Results of the discrimination testing were initially plotted as psychometric functions; that is, percent correct discrimination was plotted as a function of the number of steps in the STEP signal. For economy of space and to highlight the importance of step duration, these original psychometric functions were re-plotted. The duration of a single step (rather than number of steps in the whole signal) was chosen as the abscissa. Plotted this way, discrimination of GLIDE versus STEP signals could be compared for different signal durations. To facilitate such comparisons, three sets of psychometric functions were produced for each center frequency tested. Each set contained psychometric functions for combinations of T and ΔF which result in the same transition rate. Each set represented discrimination performance in percent correct as a function of individual step size for transition rates of 2, 4 or 8 Hz/msec.

Temporal resolution at 1 kHz

Figure 2 shows the averaged results for four listeners, for STEP vs GLIDE discrimination at $f_c = 1$ kHz. Within each panel, either two or three psychometric functions are shown. In the top panel, performance for ΔF transitions of 200 Hz over 100 msec and 100 Hz over 50 msec are displayed. In the center panel, functions for transitions of 400 Hz over 100 msec, 200 Hz over 50 msec and 100 Hz over 25 msec are plotted. The bottom panel shows 400 Hz over 50 msec and 200 Hz over 25 msec. Thus, transition rates from the top to the bottom panel of the figure are 2, 4 and 8 Hz/msec. Symbol type (open circles = 100 msec, filled circles = 50 msec and triangles = 25 msec) always indicates the duration of the signal pair. Solid lines denote $\Delta F = 400$ Hz, medium dashed lines indicate $\Delta F = 200$ Hz and dotted lines represent $\Delta F = 100$ Hz.

Figure 2 about here

Temporal Acuity with FM Glides

Visual inspection of the psychometric functions in Figure 2 shows a STEP vs GLIDE discrimination threshold of about 8 msec at 2 Hz/msec determined from the intercept at 75%. For 4 Hz/msec the value is about 7 msec, and at the highest rate, 8 Hz/msec, it is 5 msec. Note that the psychometric functions for 25- and 50-msec transitions are nearly congruent. Those for 100-msec transition rates appear to be shifted to the left, indicating somewhat better discrimination for the longer sweeps.

Temporal resolution for f_c from 250 Hz to 6 kHz

Figure 3 displays STEP vs GLIDE discrimination thresholds for center frequencies ranging from 250 Hz to 6 kHz. Discrimination threshold is taken as the step duration for $P(C) = 75\%$. The parameter is transition rate. Thus results have been collapsed over signal duration. The same four listeners participated in the experiment through 4 kHz. Four new listeners replaced the original listeners for the 6 kHz condition, which was tested several months after the original data were collected.

Figure 3 about here

For f_c s of 250, 500 and 2000 Hz, the results are similar to those obtained at 1 kHz. At 2 kHz, the psychometric functions are shifted slightly leftward, indicating greater sensitivity, as transition rate increased from 2 to 8 Hz/msec. This improvement is not evident in results at 500 and 250 Hz. The congruence of psychometric functions for 50 and 25 msec transitions and the small shift to the left for 100 msec functions, were not as marked in these results as they were at 1 kHz.

Performance in the STEP vs GLIDE discrimination task is much poorer for center frequencies of 4 and 6 kHz. For 4 kHz, discrimination threshold exceeds 20 msec at the 2 Hz/msec rate, improving to 7 or 8 msec at 8 Hz/msec. Similar results are apparent for 6 kHz, although it should be noted that new listeners replaced the original ones. There is

a tendency for performance with shorter duration transitions to be shifted toward greater sensitivity at 6 kHz, at least for the two slower rates (i.e., 2 and 4 Hz/msec).

Effect of presentation level

To determine the effect of level on STEP vs GLIDE discrimination, we repeated the testing at 1 kHz over a range of levels. At 30 and 50 dB SL, we determined STEP vs GLIDE discriminability for all three sweep rates used in the previous testing. Four new listeners were employed for this portion of the study, but procedures were essentially the same as described above. There was one important difference in the signals used in this part of the study, however. A 17-step signal was used as the standard (GLIDE) in these tests. This was because of the presence of spectral differences between the GLIDE and STEP signals that are evident about 50 dB down from the peak of the center lobe (for an example, see Figure 2 in Madden and Feth, 1992). The STEP signal contains some energy splatter not present in the GLIDE signal due to the abrupt frequency transitions. Discrimination at 70 dB SL could have been confounded by this energy splatter artifact. Using a 17-step transition as a standard should minimize the influence of this confounding factor. We chose to substitute for the original GLIDE signal rather than introducing a low-level broad band masker to cover possible splatter.

Results are displayed in Figure 4. The height of each bar indicates the temporal threshold determined at each presentation level. Bars are coded to indicate the transition rates of 2, 4 and 8 Hz/msec. Only one sweep rate was tested at 70 dB SL. The results indicate a slight decrease in temporal threshold as level increases, but the differences are very small, on the order of one or two milliseconds. The absence of a substantial improvement in performance at 70 dB SL supports the contention that these results are not contaminated by the low-level, long-term spectral cues.

Temporal Acuity with FM Glides

Note also that the results at 50 dB SL are essentially identical to those shown at 1 kHz in Figure 2. It might be suggested that the STEP signals were distinguishable from their respective GLIDE signals in the first part of the experiment because of spectral splatter. If this were true, we would expect Figure 3, which displays results obtained with the linear GLIDE to reflect better performance than Figure 4, which displays results with the 17-step "glide". This obviously is not the case.

Figure 4 about here

Discussion

Summary of findings

The STEP vs GLIDE discrimination performance of our listeners is very consistent for f_c values from 250 through 2000 Hz. On the average, a multi-step transition is distinguishable from its linear-sweep counterpart when the individual steps are 5 to 10 msec in duration. We would like to suggest that the duration of a "just discriminable" single step is a good indicator of the width of the temporal window of the normal auditory system. However, we must first rule out some obvious alternative explanations.

First, we must consider whether the listeners are using long-term spectral differences to distinguish between the GLIDE and the STEP signals. Given the abrupt change in frequency at each step, there must be some splatter of energy in the STEP signal that is not present in the GLIDE signal. Spectra for equivalent GLIDE and STEP signals reveal only very small, non-systematic differences in the main lobe. Any significant differences in the "tails" of the spectra, where we might expect off-frequency listening to occur, are more than 50 dB down from the level of the main lobe. Given that most of our testing was conducted at 50 dB SL, and the results using 17-step standard signals in place of the linear glides, we

find the off-frequency listening explanation difficult to accept.

Another explanation of our listeners' ability to distinguish GLIDE from STEP signals might suggest that since each transition results in a frequency change, listeners may be performing a simple frequency discrimination task. Two observations lead us to reject this explanation at frequencies below 2 kHz. First, at the just-discriminable step size, the frequency differences are considerably larger than the normal DLF. For example, at 1 kHz, the duration of the just discriminable step ranges from 10 msec at the 2 Hz/msec rate, to 5 msec at 8 Hz/msec. Concomitant frequency changes range from 20 Hz to 40 Hz at each step. Even considering the effect of shorter durations, these values are larger than expected from simple frequency discrimination measures (Moore, 1973). It is also difficult to understand why the just discriminable frequency change should grow from 20- to 40 Hz as the rate of transition increases from 2- to 8 Hz/msec. Further evidence against a simple frequency DL explanation lies in the almost constant performance from 250 through 2000 Hz. Discrimination dependent on frequency differences should vary with f_c as the DLF does.

However, at 4 kHz and above, listener performance may be limited by frequency discrimination. Unlike the lower frequencies, at 4 kHz the just discriminable step duration varies with transition rate. As transition rate increases from 2 Hz/msec to 8 Hz/msec, the just distinguishable duration decreases from above 20 msec to about 7 msec. If we calculate the size of frequency transition at each 75% point on the various psychometric functions, they lie in the 50 Hz range. Thus, it appears that a constant frequency jump may account for performance at 4 kHz, rather than a constant step duration.

We might then expect that the results for 6 kHz should show similar behavior, with a just discriminable frequency step somewhat larger than that at 4 kHz. If we assume that $\Delta F / F$ is constant, then at 4 kHz, $50 / 4000 = 0.0125$. At 6 kHz, the DLF should be 0.0125×6000 , or 75 Hz. While our listeners approach that value for 8 Hz/msec, most of the fre-

quency jump values at 75% performance at the slower transition rates are smaller than this predicted value. Comparison is hindered by the fact that a different set of subjects were tested at 6 kHz. Nevertheless, their performance does not show a constant just-discriminable frequency jump across transition rates at 6 kHz.

Relationship with critical bandwidth

The differential effect of bandwidth on many psychoacoustic phenomena is well known (e.g., Scharf, 1970), and the concept of the critical band, or auditory filter, has been shown to be important in the explanation of temporal resolution phenomena. For example, using pairs of sinusoidal markers of the same or different frequency, Formby and Forrest (1991) found that gap thresholds increase as the frequency separation between markers is increased. They then fit their data using a model of the auditory filter. They assumed that in the gap detection task the subject monitors the output level of a single auditory filter centered on the first marker of the marker pair. Using a roex model of the weighting function for the auditory filter, they then calculated the amount of attenuation of the second marker at various frequency separations. They were able to accurately predict the increase in gap detection threshold, due to the attenuation of the second marker, as the frequency separation between the markers increased.

We wished to see if the results from the present study can be reconciled with a model that involves monitoring the output level of the auditory filters. Greenwood (1991) has recently summarized a large body of bandwidth estimates. At 250 Hz, a good estimate of the equivalent rectangular bandwidth, ERB, is 50 Hz. All of the sweep widths used in the present study (100, 200 and 400 Hz) exceed this ERB. As f_c is increased, the smaller ΔF values approximate the ERB. At 1 kHz, for example, the ERB is about 150 Hz. Thus, the 100 Hz sweep falls within one ERB, the 200 Hz sweep just exceed one ERB, and the 400 Hz sweep traverses several bandwidths. Above 2 kHz, all sweeps are contained

within one ERB.

A 4-step signal of $f_c = 250$ Hz with a 200 Hz transition would excite those filters with center frequencies at each of its "steady-state" frequencies (approximately 150, 217, 284 and 350 Hz). Filters with center frequencies between these frequencies would be excited to a lesser extent. The GLIDE signal would, however, excite all filters between 150 and 350 Hz. Thus, a mechanism monitoring the output level of the individual filters would see two distinct excitation patterns over time: the GLIDE would excite all filters over its range equally, whereas the STEP would excite primarily those filters at its individual step frequencies, and not those skipped over by the frequency jumps. At 2000 Hz, the picture would obviously be quite different. For a 200 Hz transition, because of the increased filter width, no filters would be "skipped" in the jumps between steps; all auditory filters within the transition range would be excited to the same extent. The monitoring mechanism would see GLIDE and STEP signal filter output patterns that are much less distinct from one another than those produced at lower f_c s. Such a model would thus lead us to predict systematically poorer discrimination performance as f_c increases. However, for our results, this is clearly not the case. Thus, our results suggest that a detection system using the output levels of the auditory filters, such as has been shown to account for the detection of amplitude changes in spectrally static signals, may not be useful in explaining the temporal resolution of frequency-modulated signals.

Relationship of results to other temporal acuity measures

Estimates of auditory temporal acuity range from less than 1 msec to more than 20 msec, depending on the task used to determine the temporal threshold (see for example, Green, 1971, 1973, 1985). Our results appear to most closely resemble those from Plack and Moore's (1990) careful determination of the shape of the temporal window. Both studies produced indications of temporal resolution in the 7 to 10 msec range over much of the auditory spectrum. In the Plack and Moore study, win-

dow shapes were determined using a tone pulse located in a temporal gap between two maskers. That is, they determined the limits of temporal resolution that are imposed by forward and backward masking in a paradigm that is the temporal analog of Patterson's auditory filter shape determination (1976). The similarities between our results and theirs suggests that temporal masking may be the main limiting factor in the detection of frequency changes in the step signal. As Plack and Moore (1990) suggest, it may be that some mechanism smoothes or integrates neural activity, or neural information as they put it, over a certain time period. However, there are several interesting differences between the Plack and Moore results and those of the present study.

One difference is seen in the effect of presentation level. The Plack and Moore results show improvement in temporal resolution, i.e., a smaller ERD (equivalent rectangular duration), with increased level at all test frequencies except 300 Hz. Our comparison of performance at 30 dB SL with that at 50 dB SL shows little difference in performance. Even the limited testing at 70 dB SL shows little change.

Both studies demonstrated an effect of signal frequency, but at opposite ends of the spectrum. Plack and Moore report poorer performance at their lowest test frequency (300 Hz) but we have found degraded performance at frequencies beyond 2 kHz. Looking first at Plack and Moore's 300 Hz results, we note that estimates of auditory filter width at low frequencies have been confounded by difficulties in specifying the level of the masking noise (Fastl and Schorer, 1986). An under-specification of masker level in the low frequencies could lead to inflated estimates of masking ability whether the task is used to determine filter bandwidth or temporal window shape.

Relationship with neural synchrony

Next we consider our results for f_c above 2 kHz. The original subjects' discrimination thresholds at 4 kHz are twice those at 2 kHz for the 2 Hz/msec transition rate. Our quick check with new listeners at 6

kHz confirms this poorer performance at higher frequencies. As we suggested in the discussion above, we cannot rule out frequency discrimination, rather than temporal acuity, as the controlling factor for f_c above 2 kHz.

This poor high frequency performance contrasts with Plack and Moore's results and with results obtained in gap detection studies using deterministic signals (e.g., Formby and Forest, 1991). We are tempted to explain this difference by suggesting that our listeners' ability to perform our discrimination task is related to synchronization of the auditory nerve fiber responses. Since Plack and Moore's study required only the detection of a tone in a temporal gap between noise maskers, we should not expect their results to exhibit a dependence on synchronization. In most mammalian ears, synchrony falters above 2 kHz and is completely absent above 5 kHz (Rose, Brugge, Anderson and Hind, 1967; Anderson, Rose, Hind and Brugge, 1971). Also, Sinex and Geisler (1981) have shown that the integrity of temporal coding at the lowest instantaneous frequencies is preserved even at extremely high sweep rates, well above those used for the linear glide stimuli of the present study. They studied single-unit responses for transition rates from .2 kHz/sec for fibers with characteristic frequencies, CF, below 3 kHz and from 2 kHz/sec to 160 kHz/sec for fibers with CF above 3 kHz. (Note that these rates are reported in kHz/sec while we used Hz/msec, making them numerically equivalent.) Sinex and Geisler used a tone that followed a trapezoidal frequency trajectory with center frequency near that of the fiber's CF. Displays of inter-stimulus interval histograms show that temporal discharge patterns for the units tracked instantaneous frequency up to 2 kHz. The display is characterized as "less clear" for frequencies above 2 kHz. The neural synchrony data suggest that the mammalian auditory system is capable of temporally following the frequency modulations of the STEP - GLIDE signals at the lower frequencies but not at the higher frequencies. These findings are consistent with the results of this study.

If it is true that STEP vs GLIDE discrimination is dependent upon the

Temporal Acuity with FM Glides

synchrony of auditory nerve fibers, then one problem remains to be explained. In the 4 kHz results, discrimination improves with increased transition rates. The just discriminable step reaches 20 msec at 2 Hz/msec, but is nearer 10 msec at 8 Hz/msec. If discrimination depends upon the ability of nerve fibers to remain phase-locked to the frequency-modulated signal, we expect that just the opposite result should hold.

Conclusions

We have devised a means of determining the temporal acuity of the human auditory system using frequency modulated signals. These signals have some characteristics in common with the formant transitions of speech, and thus may be useful in relating psychoacoustic performance in this task to speech processing.

Listeners with normal hearing were asked to discriminate between sinusoids that were modulated linearly over a brief time interval from signals covering the same frequency excursion in a multiple-step trajectory. When the GLIDE signal and the STEP signal are just discriminable (75% correct in 2AFC), we assume that the duration of a single step is just less than the width of the temporal window.

For presentation frequencies from 250 Hz through 2000 Hz, our estimate of the temporal window is 7 to 10 msec. Presentation level had no effect on results, at least for 30 and 50 SL. We were cautious in our testing above 50 SL because spectral differences between test signals might confound our results.

Above 2 kHz, performance was poorer in the task. In fact, at 4 kHz we cannot rule out the possibility that our listeners were basing their decisions on just discriminable frequency jumps, rather than on temporal differences. A follow up at 6 kHz using different listeners led to equivocal results.

We are tempted to conclude that the ability of our listeners to distinguish between GLIDE and STEP signals at 2 kHz and below, is related to

the synchronization of single unit fibers in the periphery. Evidence for this conclusion could only be "circumstantial" in psychoacoustic studies. A physiological test in laboratory animals would likely be required to provide support for this speculation.

Acknowledgment

The authors wish to express their appreciation to Mary Neill, who collected most of the data presented here, and Lisa Stover, who helped to design the original experiment. Rick Gerren and Chien-yeh Hsu developed computer programs for the synthesis of the test signals and for control of the experiment. We wish to thank Gail Whitelaw, Ina Bicknell and William Melnick for their comments on a draft of this manuscript. This work was supported by a grant from the Air Force Office of Scientific Research.(89-0227).

Figure Captions

Figure 1. Simulation of auditory filter bank response to the frequency-modulated signals used in the present experiment. Each sinusoid traverses 400 Hz in 100 msec. Center frequency is 1000 Hz. The top panel shows the response for the linear glide. The next panel shows the response for a two-step transition. Each succeeding panel shows the response for 3-, 5-, 9- and 17-step transitions.

Figure 2. Each panel displays averaged performance for the four listeners as psychometric functions. The ordinate shows percent correct in the 2Q, 2AFC task. The abscissa is the duration of an individual step for the multiple-step transition. Each function within a panel represents performance for GLIDE vs STEP signals of different durations with common individual step durations. Signal duration is indicated by symbol type: open circles = 100 msec, filled circles = 50 msec and triangles = 25 msec. Width of frequency transition is shown by line type. Solid lines indicate $\Delta F = 400$ Hz, medium dashed lines = 200 Hz and dotted lines = 100 Hz. The top panel contains psychometric functions for a 2 Hz/msec transition rate. The middle panel shows performance for 4 Hz/msec; the bottom panel displays 8 Hz/msec performance.

Figure 3. Discrimination thresholds obtained from averaged psychometric functions for signal frequencies from 250 Hz to 6 kHz. The threshold is defined as the step duration that would lead to 75% correct discriminations in the 2Q, 2AFC task. Circles represent transition threshold for the 2 Hz/msec transition rate, triangles represent 4 Hz/msec, and the squares represent 8 Hz/msec. The original four listeners are represented by filled symbols for frequencies up to 4 kHz. A different group of four listeners, tested several months later, are represented by open symbols at 6 kHz.

Figure 4. The effect of presentation level on GLIDE vs STEP discrimination

Temporal Acuity with FM Glides

thresholds. At each presentation level, the bars are coded to reflect transition rate. Only one rate was tested at 70 dB SL.

Temporal Acuity with FM Glides

List of References

- Anderson, D.J., Rose, J. E., Hind, J. E. and Brugge, J. F. (1971). "Temporal position of discharges in single auditory-nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects," J. Acoust. Soc. Am. 49, 1131-1139.
- Buus, S., and Florentine M. (1985). "Gap detection in normal and impaired listeners: the effect of level and frequency," in Time Resolution in Auditory Systems, edited by A. Michelsen (Springer, Berlin).
- Buunen, T.J.F., and Valdenburg, D.A. van (1979). "Auditory detection of a single gap in noise," J. Acoust. Soc. Am. 65, 534-537.
- Eddins, D. A., Hall, III, J. W. and Grose, J. H. (1992). "The detection of temporal gaps as a function of frequency region and absolute noise bandwidth," J. Acoust. Soc. Am. 91, 1069-1077.
- Fastl, H. and Schorer, E. (1986). "Critical bandwidth at low frequencies reconsidered," in Auditory Frequency Selectivity, edited by B. C. J. Moore and R. D. Patterson (Plenum Press, New York and London).
- Fitzgibbons, P.J. (1979). "Temporal resolution in normal and hearing-impaired listeners," Ph.D. Dissertation, Northwestern University.
- Fitzgibbons, P.J., and Wightman, F.L. (1982). "Gap detection in normal and hearing-impaired listeners," J. Acoust. Soc. Am., 72, 761-765.
- Florentine, M., and Buus, S. (1984). "Temporal gap detection in sensorineural and simulated hearing impairments," J. Sp. Hear. Res. 27, 449-455.
- Formby, C., and Forrest, T.G. (1991). "Detection of silent temporal gaps in sinusoidal markers," J. Acoust. Soc. Am. 89, 830-837.
- Geisler, C.D. and Sinex, D.G. (1982). "Responses of primary auditory fibers to brief tone bursts," J. Acoust. Soc. Am., 72, 781-794.
- Glasberg, B.R., Moore, B.C.J. and Bacon, S.P. (1987). "Gap detection and masking in hearing-impaired and normal hearing subjects," J. Acoust. Soc. Am., 81, 1546-1556.
- Green, D. M. (1971). "Temporal auditory acuity," Psych. Rev. 78, 540-551

Temporal Acuity with FM Glides

Green, D. M. (1973). "Minimum integration time," in Basic Mechanism in Hearing, edited by A. Moller, (Academic Press, New York), pp 829-846.

Green, D.M. (1985). "Temporal factors in psychoacoustics," in Time Resolution in Auditory Systems, edited by A. Michelsen (Springer, Berlin).

Green, D.M., and Forrest, T.G. (1988). "Detection of amplitude modulation and gaps in noise," in Basic Issues in Hearing, edited by H. Duifhuis, J.W. Horst, and H.P. Wit (Academic Press, New York).

Greenwood, D. D. (1991). "Critical bandwidth and consonance in relation to cochlear frequency-position coordinates," Hear. Res. **54**, 164-208.

Huffman, D.A. (1962). "The generation of impulse-equivalent pulse trains," IRE Transactions, IT8, 510-516.

Kay, R.H. (1982). "Hearing of modulation in sounds," Physiological Reviews, **62**, 894-975.

Madden, J. P. and Feth, L. L. (1992). "Temporal resolution in Normal-hearing and hearing-impaired listeners using frequency-modulated stimuli," J. Speech and Hearing Research **35**, 436-442.

Moore, B.C.J. (1973). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610-619.

Moore, B.C.J. (1989). An Introduction to the Psychology of Hearing, 3rd Edition. (Academic Press, London).

Moore, B.C.J., and Glasberg, B.R. (1988). "Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears," J. Acoust. Soc. Am. **83**, 1093-1101.

Moore, B.C.J., Glasberg, B.R., Donaldson, E., McPherson, T. and Plack, C. (1989). "Detection of temporal gaps in sinusoids by normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **85**, 1266-1275.

Moore, B.C.J., Glasberg, B.R., Plack, C.J. and Biswas, A.K. (1988). "The shape of the ear's temporal window," J. Acoust. Soc. Am. **83**, 1102-1116.

Patterson, R. D. (1976). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640-654.

Temporal Acuity with FM Glides

- Patterson, R. D. and Moore, B. C. J. (1986). "Auditory filters and excitation patterns as representations of frequency resolution," in Frequency Selectivity in Hearing, edited by B. C. J. Moore, (Academic Press, London and New York).
- Patterson, J.H. and Green, D. M. (1970). "Discrimination of transient signals having identical energy spectra." J. Acoust. Soc. Am. 48, 984-905.
- Patterson, R. D., Robinson, K., Holdsworth, J., Mckeown, D., Zhang, C., and Allerhand, M. (1992). "Complex sounds and auditory images," in Auditory Physiology and Perception, edited by Y. Cazals, L. Demany and K. Horner, (Pergamon Press, Oxford)
- Penner, M.J. (1975) . "Persistence and integration: Two consequences of a sliding integrator," Percept. Psychophys. 18, 114-120.
- Penner, M.J. (1977). "Detection of temporal gaps in noise as a measure of the decay of auditory sensation," J. Acoust. Soc. Am. 61, 552-557.
- Penner, M. J., and Cudahy, E. (1973). "Critical masking interval: A temporal analog of the critical band," J. Acoust. Soc. Am. 65, 1530-1534.
- Penner, M. J., Cudahy, E., and Jenkins, G. W. (1974). "The effect of masker duration on forward and backward masking," Percept. Psychophys. 15, 405-410.
- Penner, M. J., Robinson, C. E., and Green, D. M. (1972). "The critical masking interval," J. Acoust. Soc. Am. 52, 1661-1668.
- Plack, C.J., and Moore, B. C. J. (1990). "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. 87, 2178-2187.
- Plomp, R. (1964). "Rate of decay of auditory sensation," J. Acoust. Soc. Am. 36, 277-282.
- Rose, J. E., Brugge, J.F., Anderson, D. J., and Hind, J. E. (1967). "Phase-locked response to low-frequency tones in single auditory-nerve fibers of the squirrel monkey," J. Neurophysiol. 34, 769-793.
- Robinson, C. E., and Pollack, I. (1973). "Interaction between forward and backward masking: A measure of the integrating period of the auditory system," J. Acoust. Soc. Am. 53, 1313-1316.

Temporal Acuity with FM Glides

Scharf, B. (1970). "Critical bands" in Foundations of Modern Auditory Theory Vol. I, edited by J. V. Tobias (Academic Press, New York).

Scott, D. (1986). "Measurement of temporal resolution in the auditory system," Doctoral thesis, Vanderbilt University.

Shailer, M.J., and Moore, B.C.J. (1983). "Gap detection as a function of frequency, bandwidth, and level," J. Acoust. Soc. Am. 74, 467-473.

Shailer, M.J., and Moore, B.C.J. (1985). "Detection of temporal gaps in bandlimited noise: effects of variations in bandwidth and signal-to-noise ratio," J. Acoust. Soc. Am. 77, 635-639.

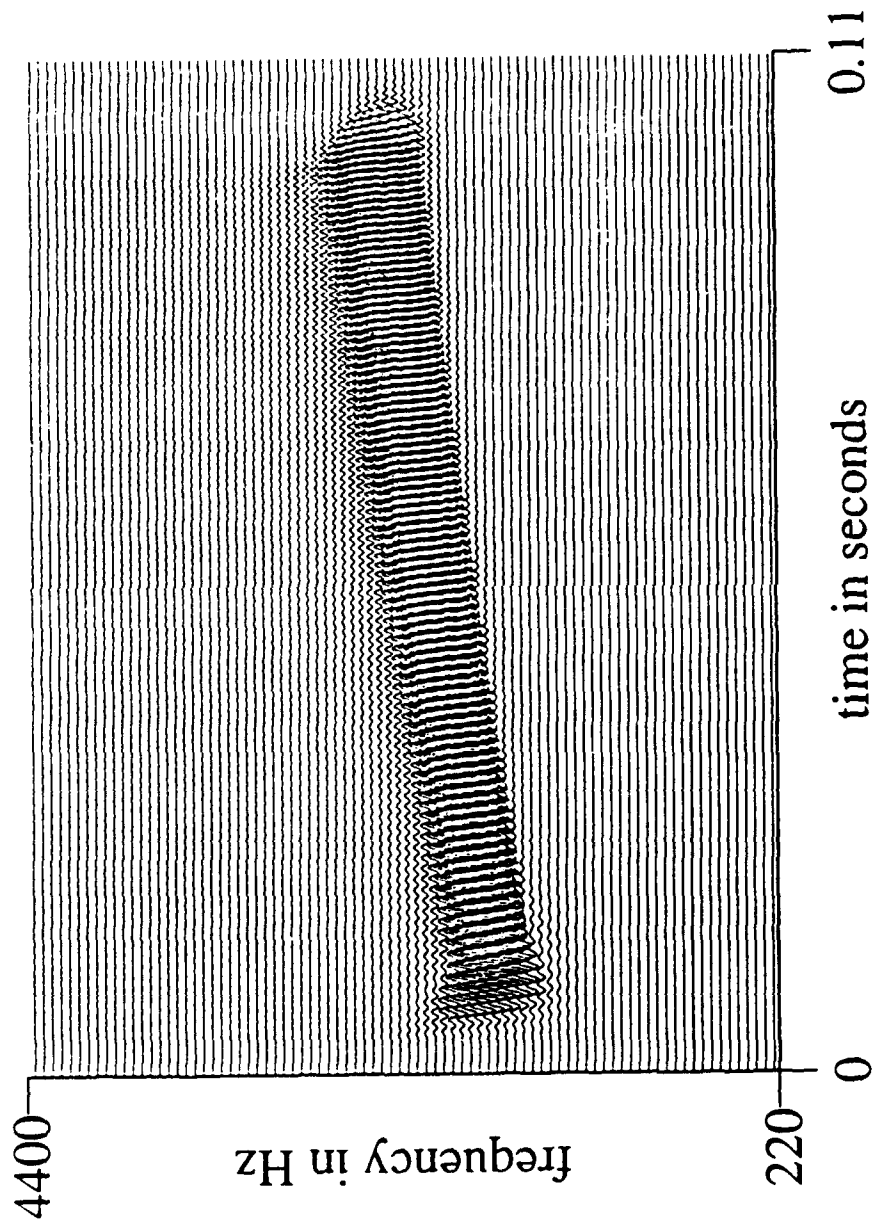
Shailer, M.J., and Moore, B.C.J. (1987). "Gap detection and the auditory filter: phase effects using sinusoidal stimuli," J. Acoust. Soc. Am. 81, 1110-1117.

Sinex, D.G., and Geisler, C.D. (1981). "Auditory-nerve fiber responses to frequency-modulated tones," Hearing Research 4, 127-148.

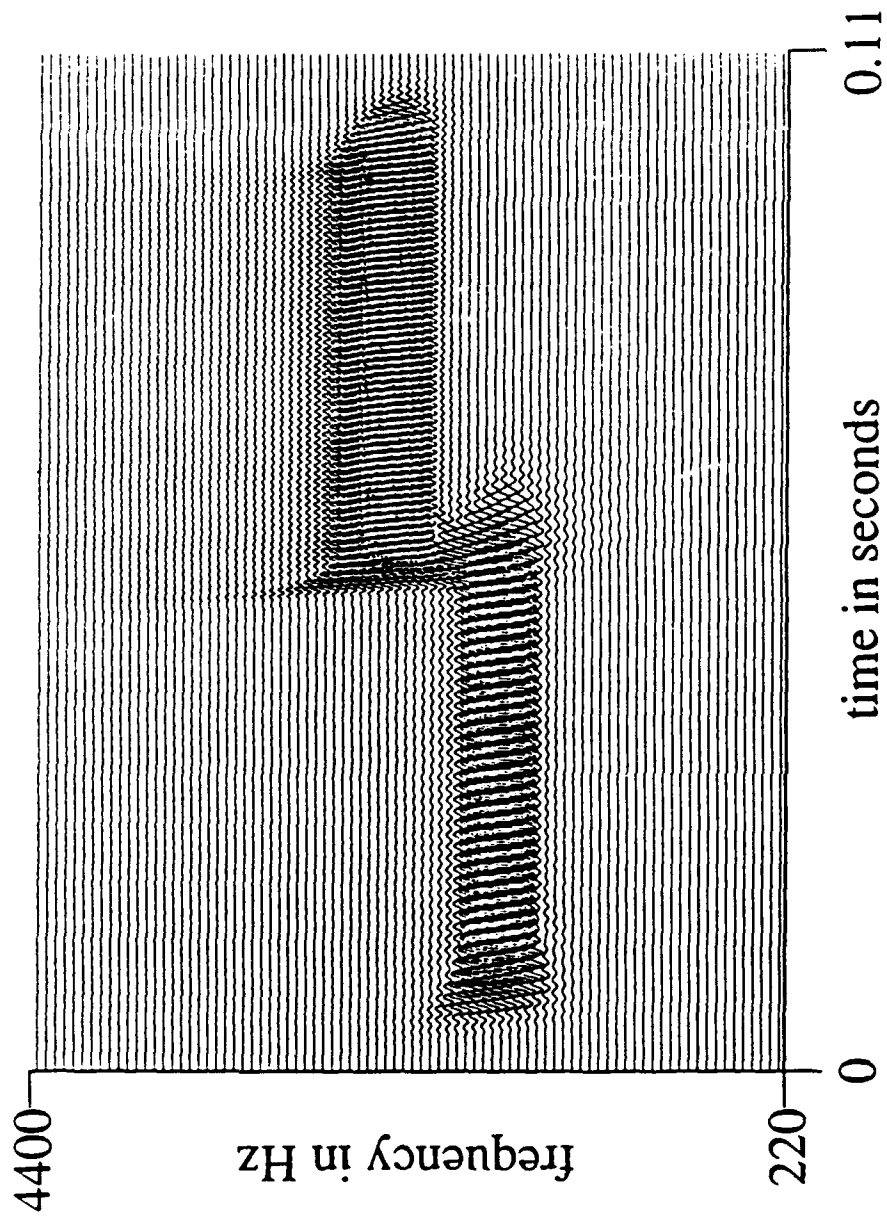
Sinex, D. G., and Geisler, C. D. (1983). "Response of auditory-nerve fibers to consonant-vowel syllables," J. Acoust. Soc. Am. 73, 602-615.

Viemeister, N.F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. 66, 1354-1380.

g4c1c.fbm



f2s4c1c.fbm



f3s4c1c.fbm

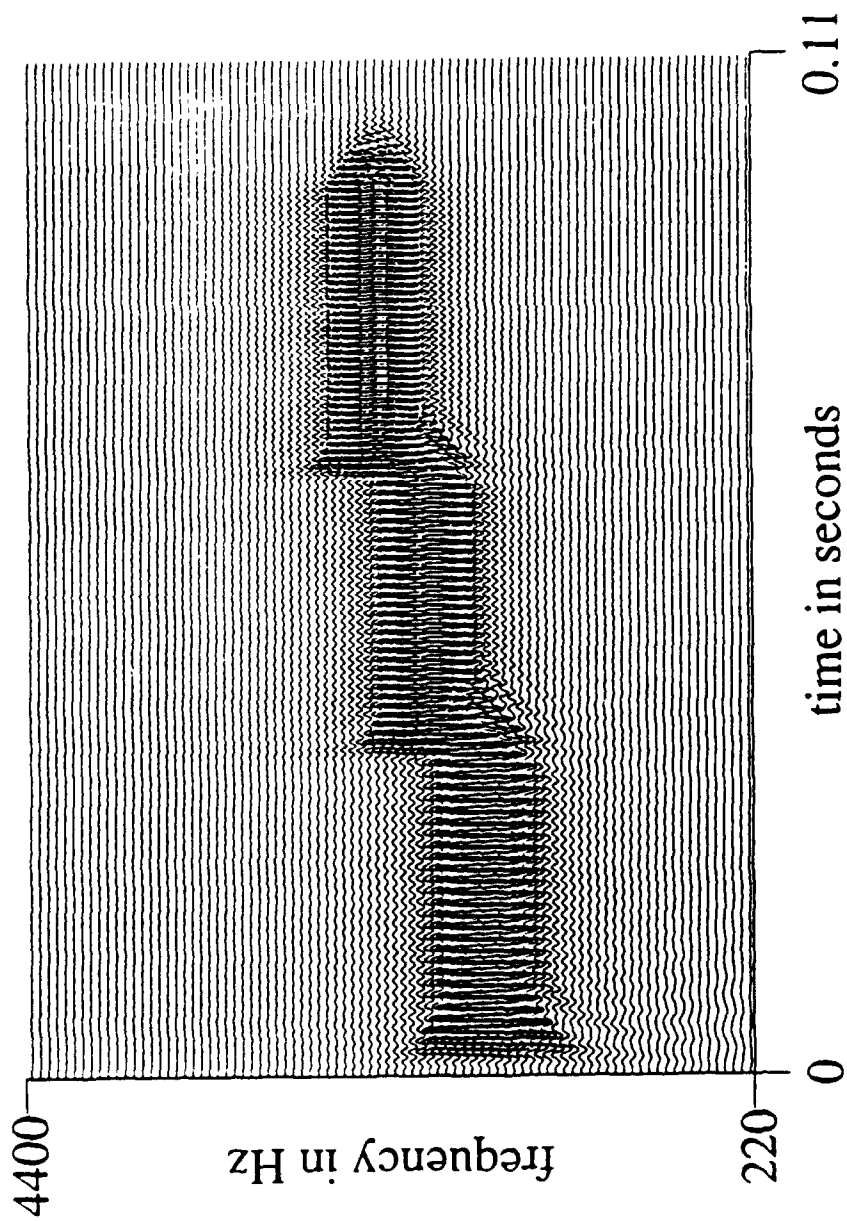


Fig 1c

f5s4c1c.fbm

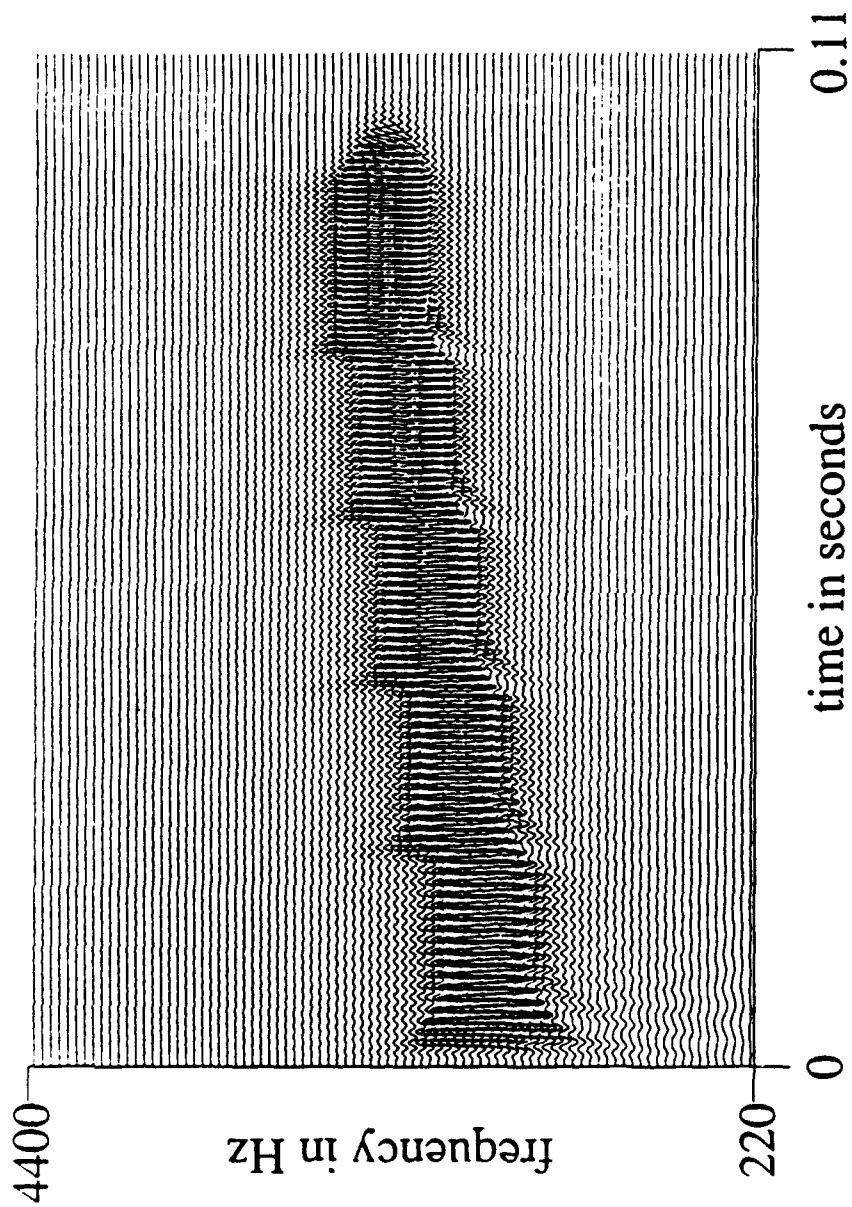


Fig 1d

f9s4c1c.fbm

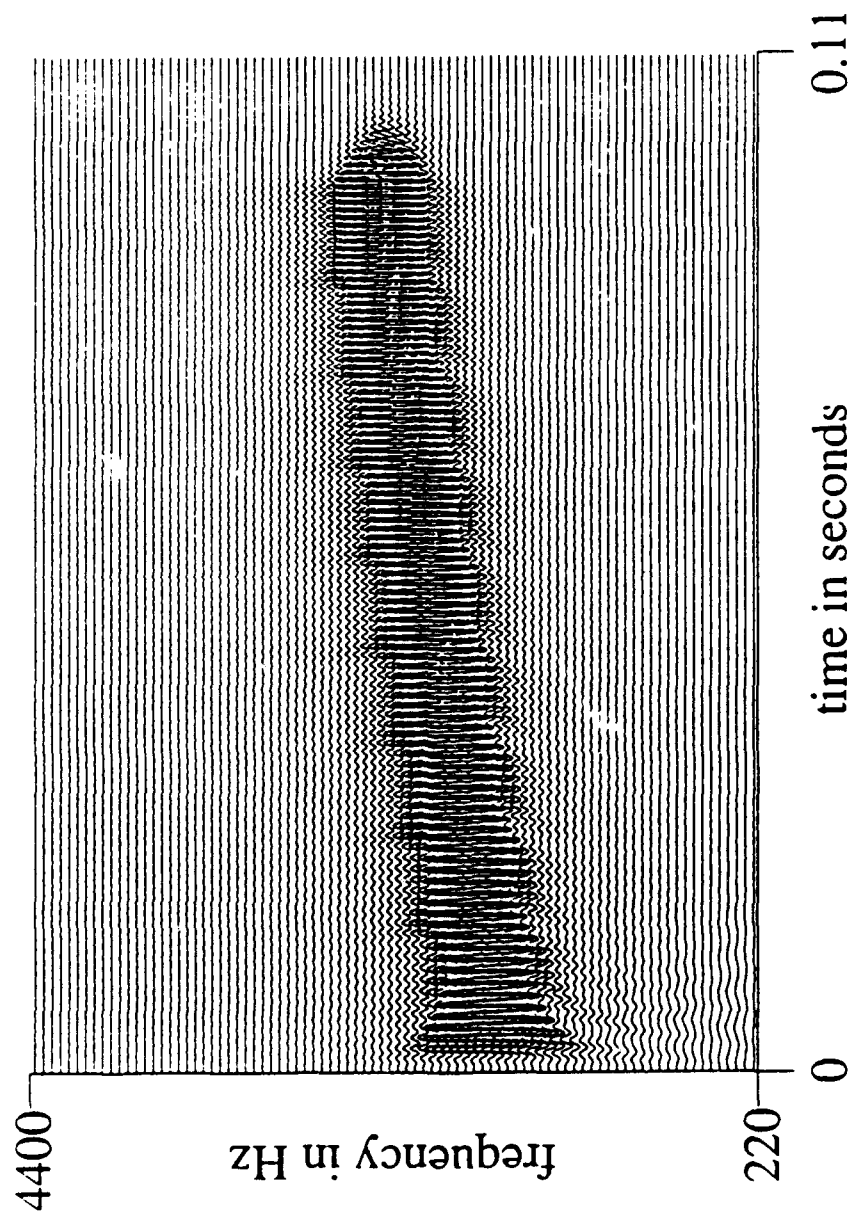
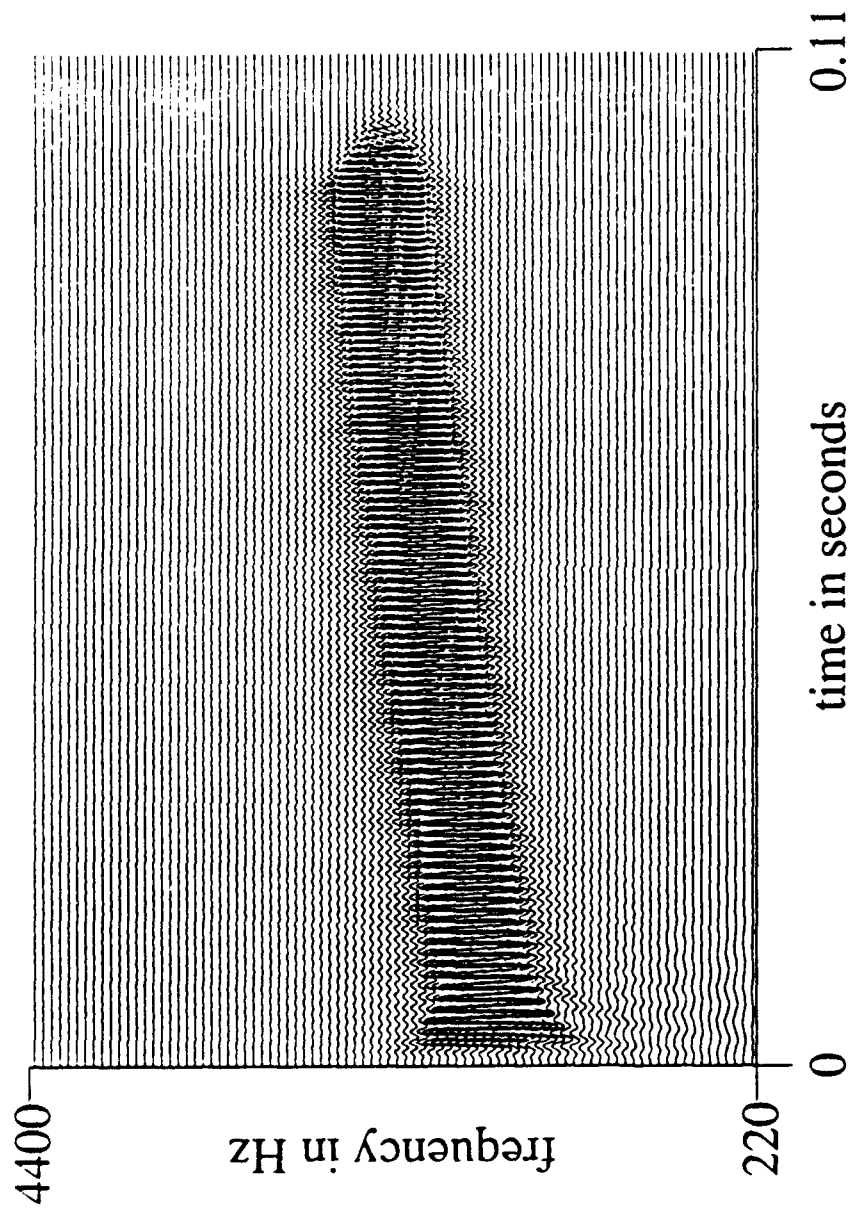


Fig 1e

f17s4c1c.fbm



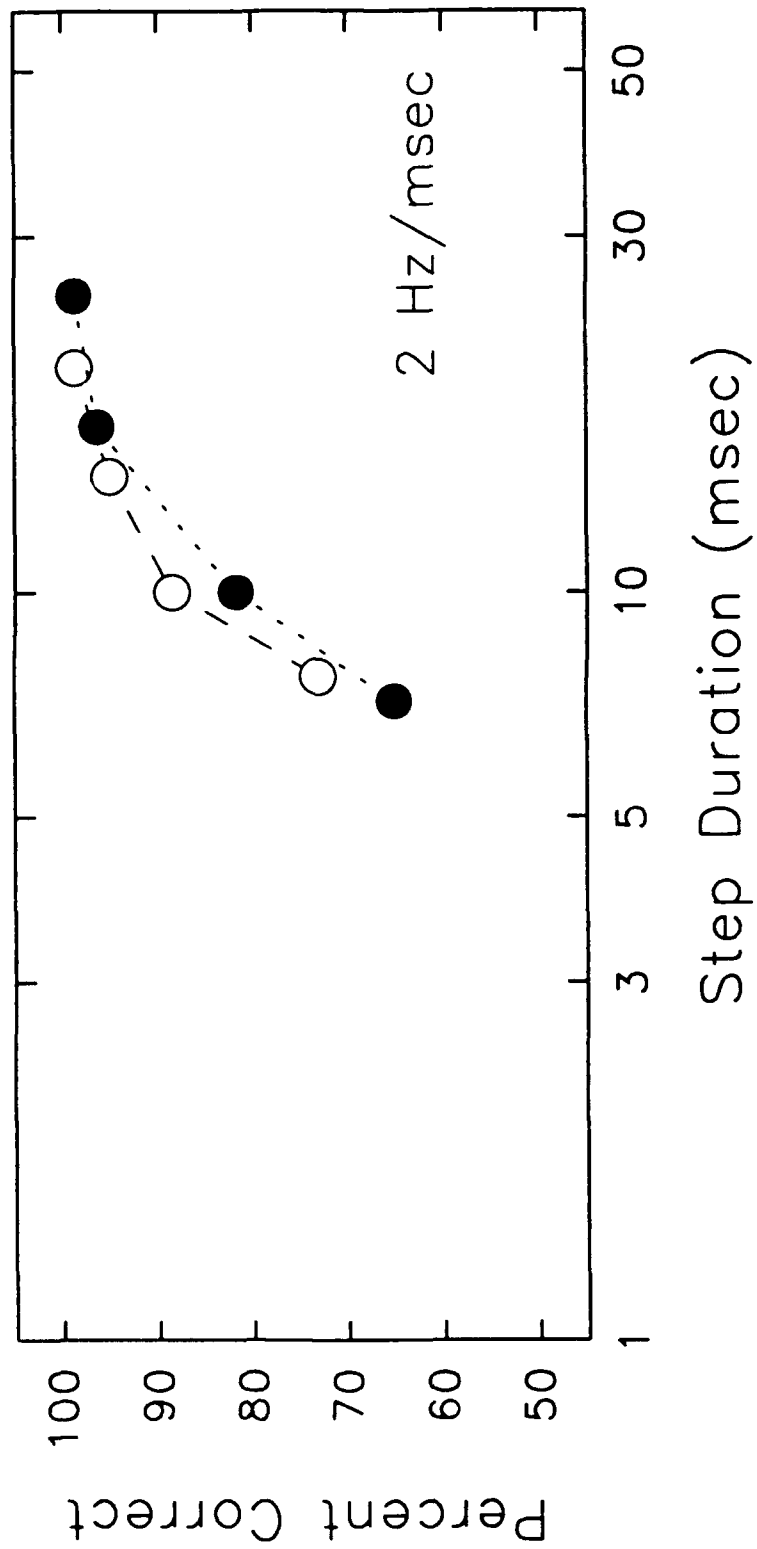


Fig 2 a

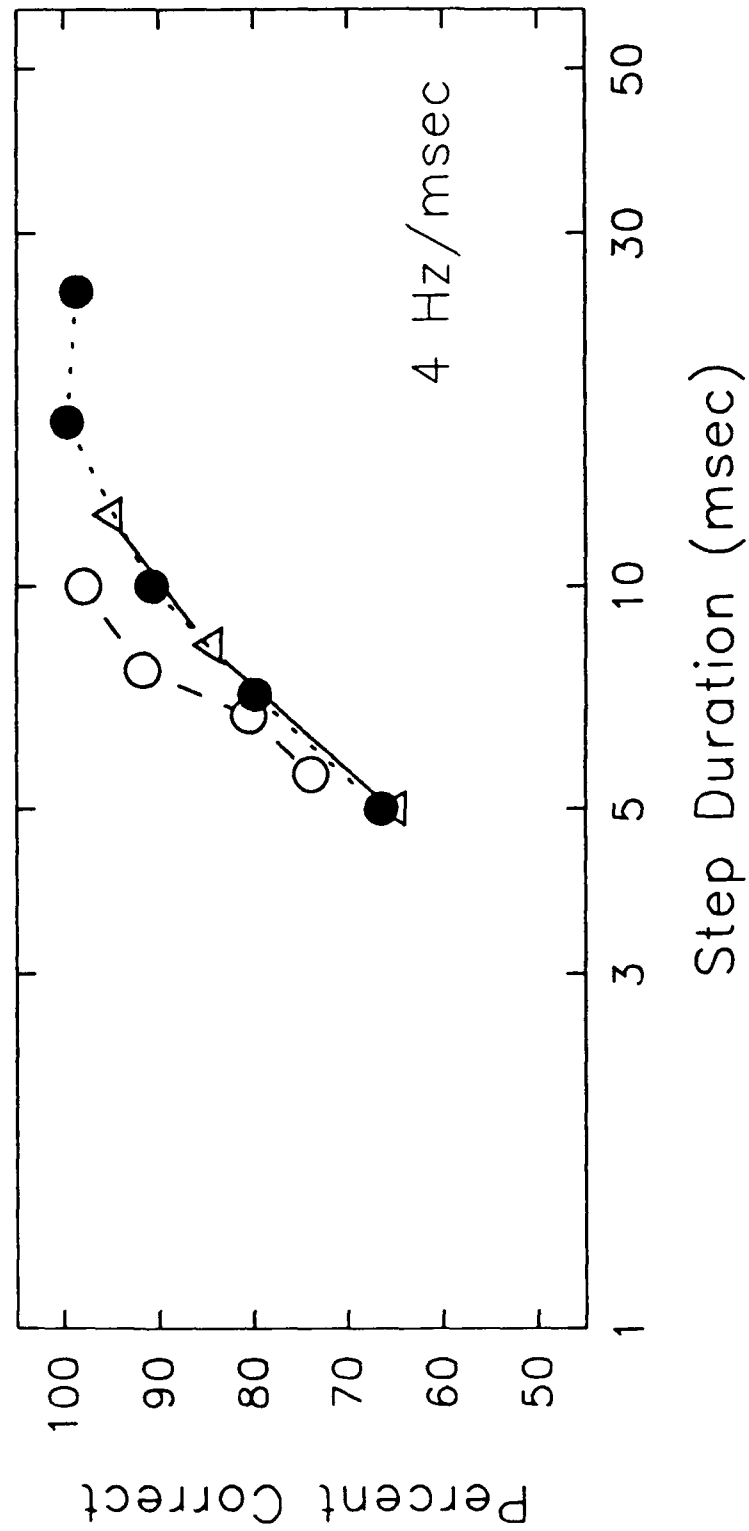


Fig 2 b

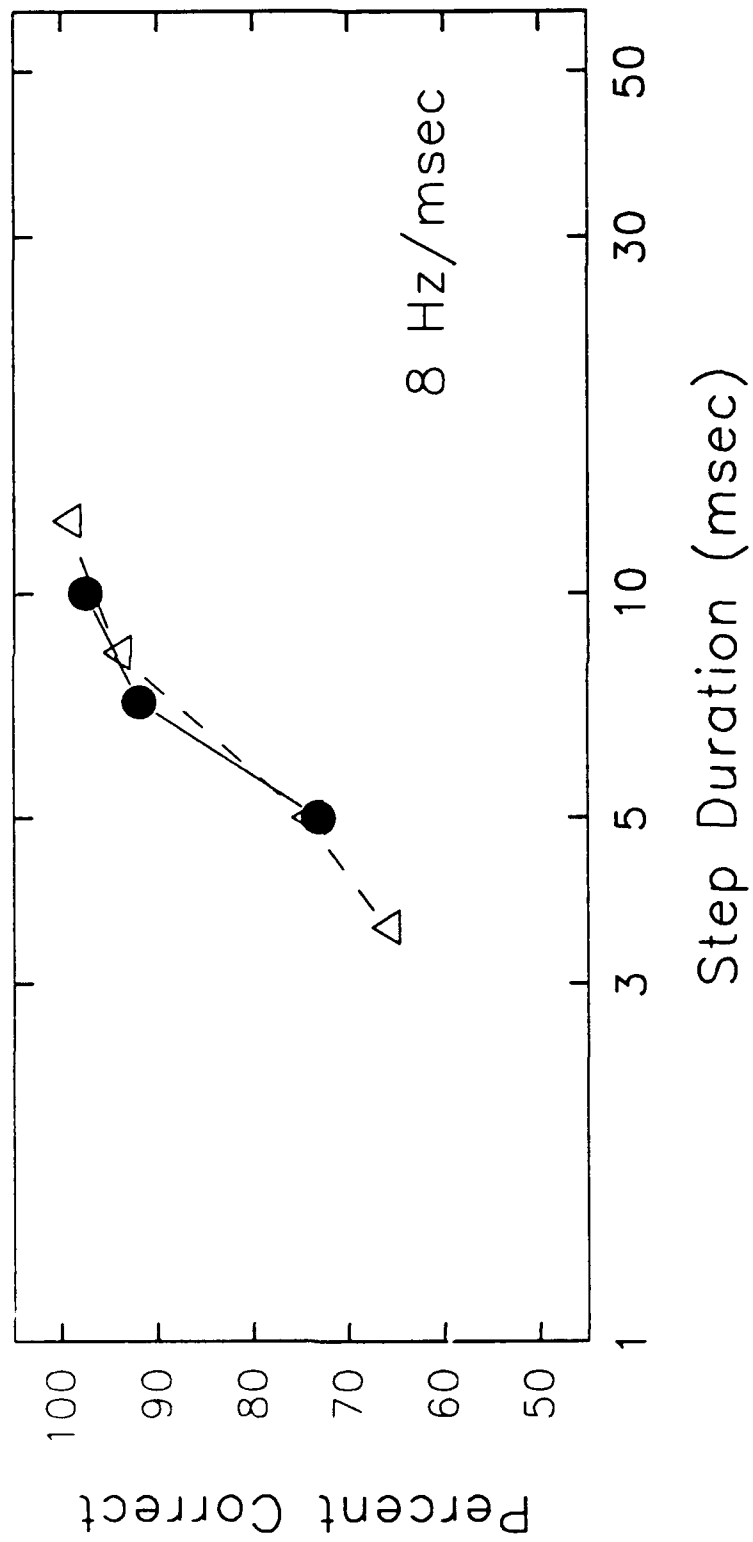


Fig 2c

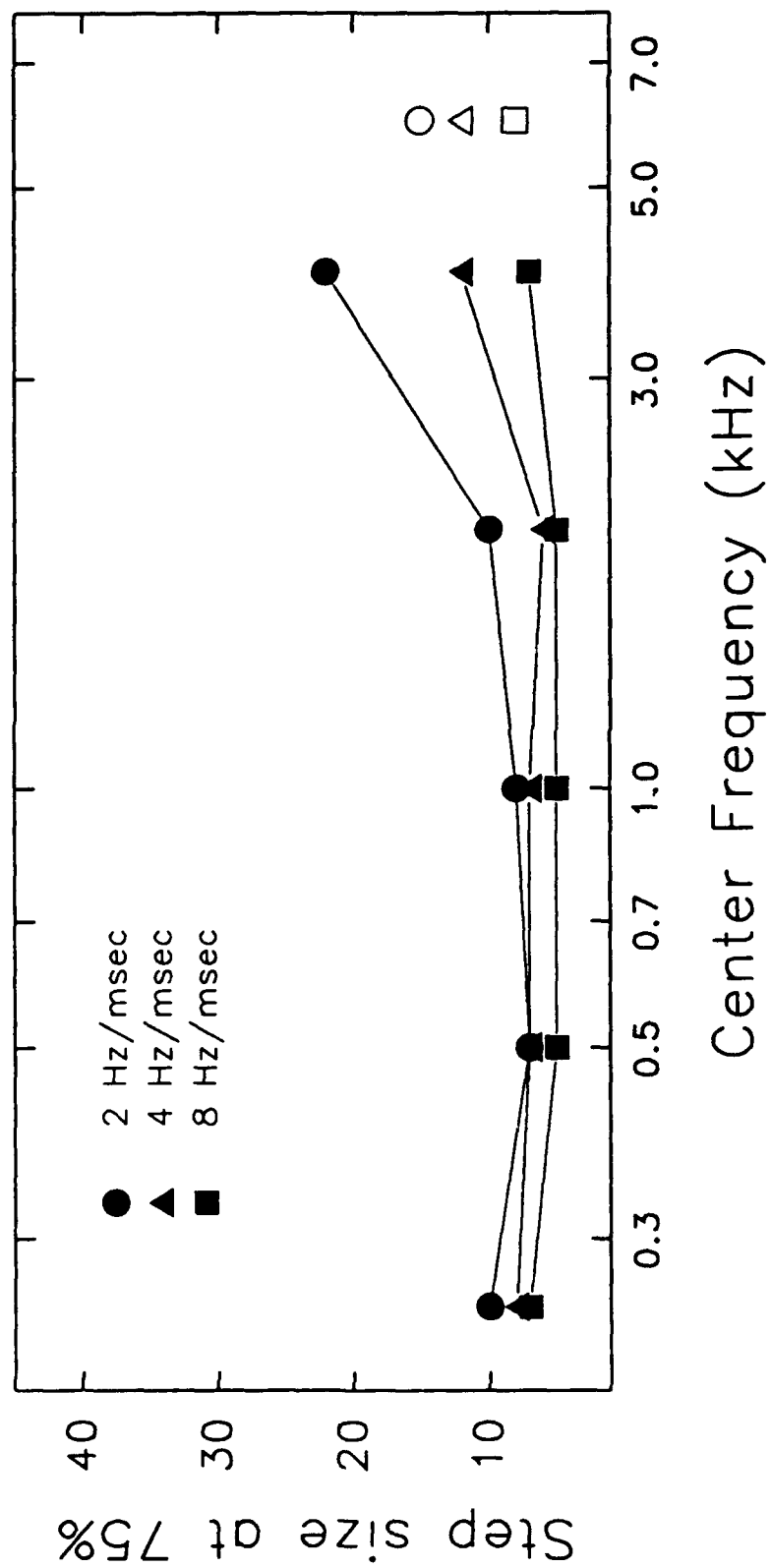


Fig 3

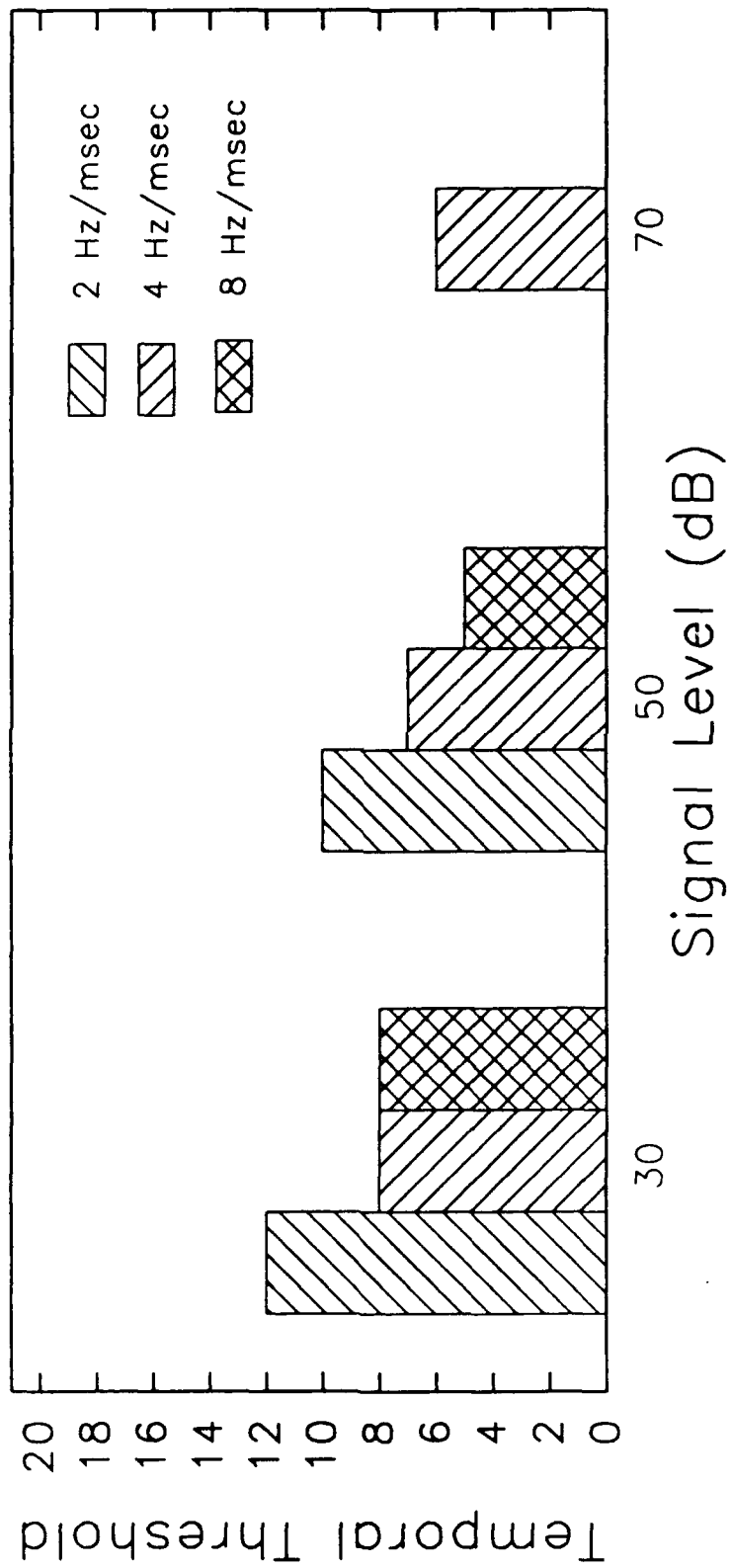


Fig 4

Intensity weighted average of instantaneous frequency as a model for frequency discrimination

Jayanth N. Anantharaman and Ashok K. Krishnamurthy

Department of Electrical Engineering

The Ohio State University

Columbus, OH 43201

Lawrence L. Feth

Department of Speech and Hearing Science

The Ohio State University

Columbus, OH 43201

PACS: 43.66.Ba, 43.66.Fe

Received _____

ABSTRACT

The intensity weighted average of instantaneous frequency (IWAIF) is developed as a model to predict listener performance in tasks primarily requiring frequency discrimination. IWAIF is closely related to the EWAIF model proposed by Feth for similar tasks. The primary difference is that the IWAIF model uses intensity (envelope-squared) as the weighting function instead of the envelope. The advantages of IWAIF over EWAIF are that (a) it has a convenient frequency domain interpretation; and (b) it is much simpler to compute than the EWAIF.

PACS numbers: 43.66.Ba, 43.66.Fe

INTRODUCTION

The envelope weighted average of instantaneous frequency (EWAIF) model was developed nearly two decades ago by Feth (1974) to account for the discriminability of two-tone complexes. Helmholtz (1954) reported that the pitch of a two-component complex tone is shifted towards the frequency of the component whose amplitude is increased slightly. Helmholtz attributed the pitch shift to fluctuations in the instantaneous frequency of the two-tone complex. Feth and coworkers (Feth, 1974; Feth and O'Malley, 1977; Feth et al., 1982) have studied the discriminability of complementary pairs of two-tone complexes (Voelcker, 1966a, b). Feth showed that the pitch differences are proportional to the EWAIF differences between the complex signals. Since then the EWAIF model has been used to explain a variety of discrimination tasks where the pitch of the stimulus is the dominant cue. For example, Feth and Stover (1987) extended the model to explain an anomaly in data relating to "profile signals" (Green, 1988). The central theme of this model is that for certain signal pairs, listeners use pitch differences to discriminate between them. Feth's model attempts to quantify the pitch changes observable in the discrimination of complex stimuli. In the case of profile signals it is assumed that changes in spectral shape of the profile signals produce a noticeable change in the perceived pitch.

Computing the EWAIF of a signal is sometimes difficult, especially for wideband signals. One problem that arises is due to the fact that the derivative of signals have

to computed in order to arrive at the EWAIF. Differentiation is a highly noise sensitive operation which may lead to incorrect values of the EWAIF. Also, it is sometimes necessary to compute the ratio of two numbers that are nearly zero. This may not be possible on computers because of the word length being finite. It has been reported that other weighting functions such as intensity (square of the envelope) perform equally well in predicting pitch differences (Feth et al., 1982). Anantharaman et al. (1991) used such a intensity weighted average of instantaneous frequency (IWAIF) model to predict frequency differences. In this paper, we shall further investigate the IWAIF model in terms of its computational difficulty and its capability to predict pitch differences. It is found that the IWAIF model is easier to compute than the EWAIF model and is also much faster. The IWAIF model does not have the drawbacks of the EWAIF model mentioned above.

First, the time and frequency domain representations of the EWAIF is presented. The IWAIF of a signal is then defined, and its representation in the frequency domain is derived. The performance of the IWAIF model is then compared to that of the EWAIF model in a number of psychoacoustic tasks.

I. EWAIF MODEL

A. EWAIF in the time domain

In general, a finite energy real signal $s(t)$ which has a Fourier transform

$$S(f) = \mathcal{F}[s(t)] = \int_{-\infty}^{\infty} s(t) e^{-2\pi j f t} dt \quad (1)$$

can be represented as (McGillem, 1979; Voelcker, 1966a, b),

$$s(t) = e(t) \cos \phi(t) \quad 0 \leq t \leq T \quad (2)$$

$$= \operatorname{Re}^1 \left[e(t) e^{j\phi(t)} \right] \quad (3)$$

where $e(t)$ is the instantaneous envelope and $\phi(t)$ is the instantaneous phase. The instantaneous frequency, $f(t)$ is defined as,

$$f(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt} \quad (4)$$

Such a representation of $s(t)$ is not unique. For example, $e(t)$ can be chosen to satisfy (3) for an arbitrary $\phi(t)$. A unique $e(t)$ and $\phi(t)$ can be assured by imposing an additional constraint, namely, that the real and imaginary parts of the complex signal $e(t) e^{j\phi(t)}$ form a Hilbert transform pair. Such a complex signal is termed

analytic and has certain useful properties. Thus, the analytic signal corresponding to the real signal $s(t)$ can be written as

$$m(t) = s(t) + j \hat{s}(t) \quad (5)$$

$$= |m(t)| e^{j\phi(t)} \quad (6)$$

where

$$\hat{s}(t) = \mathcal{H}[s(t)], \text{ the Hilbert transform of } s(t). \quad (7)$$

The envelope and instantaneous frequency functions, $e(t)$ and $f(t)$, can be defined in terms $s(t)$ and $\hat{s}(t)$ as

$$e(t) = |m(t)| = [s^2(t) + \hat{s}^2(t)]^{\frac{1}{2}} \quad (8)$$

$$\phi(t) = \arctan \left[\frac{\hat{s}(t)}{s(t)} \right] \quad (9)$$

$$f(t) = \frac{s(t) \hat{s}'(t) - s'(t) \hat{s}(t)}{s^2(t) + \hat{s}^2(t)} \quad (10)$$

The envelope weighted average of instantaneous frequency (EWAIF) of $s(t)$ is defined as

$$\text{EWAIF}[s(t)] = \frac{\int_0^T e(t) f(t) dt}{\int_0^T e(t) dt} \quad (11)$$

A common method of calculating the EWAIF of a signal is to determine the envelope and instantaneous frequency functions using (8),(10) and computing the required integrals in (11). However, there are some computational problems when we adopt this method for calculating the EWAIF of broadband signals. Note that the expression (10) for $f(t)$ involves differentiation which is a highly noise sensitive operation.

B. Frequency domain representation of EWAIF

Alternatively $f(t)$ can be expressed in terms of the analytic signal, $m(t)$, alone by rewriting (6) as

$$\ln m(t) = \ln |m(t)| + j \phi(t) \quad (12)$$

Hence,

$$\phi(t) = \text{Im}[\ln m(t)]; \quad \text{Im denotes the imaginary part operator} \quad (13)$$

$$f(t) = \frac{1}{2\pi} \text{Im} \left[\frac{m'(t)}{m(t)} \right] \quad (14)$$

Inserting the above equations in the expression for EWAIF (11) we have

$$\text{EWAIF}[s(t)] = \frac{1}{2\pi} \frac{\int_0^T |m(t)| \text{Im} \left[\frac{m'(t)}{m(t)} \right] dt}{\int_0^T |m(t)| dt} \quad (15)$$

This can be expressed in terms of the Fourier transform of $\sqrt{m(t)}$ as (see Appendix A for the derivation)

$$\text{EWAIF}[s(t)] = 2 \frac{\int_{-\infty}^{\infty} f |M_S(f)|^2 df}{\int_{-\infty}^{\infty} |M_S(f)|^2 df} \quad (16)$$

where $M_S(f) = \mathcal{F} [\sqrt{m(t)}]$.

The EWAIF is thus the frequency of the “center of gravity” of $|M_S(f)|^2$. While this is an interesting observation, it is of little use in the computation of the EWAIF of a signal. Indeed, in order to obtain $M_S(f)$, the square root of a complex signal has to be computed. In computing $\sqrt{m(t)}$, we have to be careful to choose the principal branch of the square root. This is similar to the phase unwrapping problem encountered in signal processing. Further, because $f(t)$ has an $e(t)$ term in the denominator, care must be taken in computing the instantaneous frequency at points where the envelope is zero or near zero. This involves computing a limit of the ratio of two functions which approach zero rather than a simple division.

II. IWAIF

A. IWAIF in the time domain

In computing the EWAIF, the envelope of the signal is used as the weighting function for finding the average of the instantaneous frequency. Other weighting functions may model listener discriminability as well. Indeed, Feth et al. (1982) observe,

Our previous modeling of the discriminability of two-component complex tones has concentrated on the envelope-weighted arithmetic average of the instantaneous frequency fluctuation in each complex signal. For these results we investigated the predictions of another weighting function, the instantaneous intensity (envelope-squared). Also, we calculated the root-mean-square (rms) average of the instantaneous frequency, and the rms of the envelope-weighted instantaneous frequency.

In all cases, the relations among the three model predictions do not vary with respect to one another. The prediction of the envelope-weighted arithmetic average model lies above that of the rms of envelope-weighted arithmetic average model, which in turn is above the intensity-weighted arithmetic average model. The differences among the predictions for these three models are so small that similar duplication was avoided in plotting the comparisons for subjects 2 through 4. These small differences among the three models can be in-

terpreted as support for the weighted time average of instantaneous frequency models in general. They indicate an insensitivity to the choice of the actual weighting function (envelope or intensity) and an insensitivity to the choice of the averaging (arithmetic versus rms).

Let us investigate the intensity weighted (arithmetic) average of instantaneous frequency (IWAIF) of a signal. The IWAIF of $s(t)$ is defined as

$$\text{IWAIF}[s(t)] = \frac{\int_0^T e^2(t) f(t) dt}{\int_0^T e^2(t) dt} \quad (17)$$

B. Frequency domain representation of IWAIF

Much of the discussion in this section follows that in Anantharaman (1992). We can rewrite (17) in terms of $m(t)$ as

$$\text{IWAIF}[s(t)] = \frac{\int_0^T |m(t)|^2 \text{Im} \left[\frac{m'(t)}{m(t)} \right] dt}{\int_0^T |m(t)|^2 dt} \quad (18)$$

Invoking Parseval's relation this becomes (see Appendix)

$$\text{IWAIF}[s(t)] = \frac{\int_{-\infty}^{\infty} f |M(f)|^2 df}{\int_{-\infty}^{\infty} |M(f)|^2 df} \quad (19)$$

This can be further simplified by taking advantage of the one-sided nature of $M(f)$ and its relation to $S(f)$. Equation (19) then becomes

$$\text{IWAIF}[s(t)] = \frac{\int_0^\infty f |S(f)|^2 df}{\int_0^\infty |S(f)|^2 df} \quad (20)$$

Thus, the IWAIF of a real signal is located exactly at the "center of gravity" of the positive portion of its energy density spectrum. Compare this with (16) which is the frequency domain expression for the EWAIF. Computing the IWAIF using relation (20) involves simply taking the Fourier transform of the signal. Using the Fast Fourier Transform (FFT) algorithm this can be done easily.

An alternative derivation for the IWAIF is to find a suitable f_0 such that $s(t)$ represents a modulated wave of the form

$$s(t) = e(t) \cos[2\pi f_0 t + \theta(t)] \quad (21)$$

$$= \text{Re}[\psi(t)] \quad (22)$$

$e(t)$ is thought of as the envelope of $s(t)$ and $\theta(t)$ as its phase. For narrow-band $e(t)$ and $\theta(t)$ this represents the modulation of a sinusoidal carrier wave of frequency f_0 . The instantaneous frequency of the signal is

$$f(t) = f_0 + \frac{1}{2\pi} \frac{d\theta(t)}{dt} \quad (23)$$

The choice of f_0 can be arbitrary so long as the mathematical relations remain valid. The most common choice (McGille, 1979) is to select f_0 such that it is the center of gravity of $|\Psi(f)|^2$. This corresponds to the center of gravity of the positive frequency portion of the energy density spectrum of the signal. The required value for f_0 is that value which minimizes the following integral

$$\int_0^{\infty} (f - f_0)^2 |\Psi(f)|^2 df \quad (24)$$

which is the same as the IWAIF of the signal $s(t)$.

C. Computation of IWAIF

The above frequency domain representation (20) provides a simple and efficient procedure for computing the IWAIF of a signal. It eliminates most of the difficulties encountered in computing the EWAIF of the signal. Moreover, the IWAIF is completely described by the energy spectrum of the *signal* alone. This obviates the need to compute a Hilbert transform and a derivative. All that needs to be computed is the Fourier transform of $s(t)$. This can be done efficiently using the FFT algorithm. Suppose $s(t)$ is sampled at a rate F_s to yield N samples, $s[n], n = 0, 1, \dots, N - 1$. Its N -point FFT is $S[k], k = 0, 1, \dots, N - 1$, say. Then, the IWAIF of $s(t)$ can be

computed as:

$$\text{IWAIF}[s(t)] \approx \frac{\sum_{k=0}^{N-1} k \Delta f |S(k)|^2 \Delta f}{\sum_{k=0}^{N-1} |S(k)|^2 \Delta f} \quad (25)$$

$$= \Delta f \frac{\sum_{k=0}^{N-1} k |S(k)|^2}{\sum_{k=0}^{N-1} |S(k)|^2} \quad (26)$$

where $\Delta f = F_s/N$ is the frequency spacing between samples of the FFT.

III. COMPARISON OF IWAIF AND EWAIF

As mentioned earlier, the only difference between the IWAIF of a signal and its EWAIF is in the choice of a weighting function. While the envelope is used to weight the instantaneous frequency in calculating the EWAIF, the intensity (envelope-squared) is used as the weighting function in IWAIF calculations. Since both envelope and intensity are non-negative and the latter is the square of the former, the weighting functions are highly correlated. Thus, similar values are expected for the EWAIF and IWAIF of a signal. For a simple sinusoid, both the EWAIF and the IWAIF values are equal to the tone frequency f_0 . For a combination of two tones of the same amplitude the EWAIF and IWAIF values are again equal and are located at the mean of the two frequencies. It is difficult to analytically calculate the EWAIF of a combination of tones. However, the IWAIF of an N -component complex can be easily calculated. Assuming T to be much larger than the maximum of all the tone periods, the IWAIF

of a sum of sinusoids such as

$$s(t) = \sum_i a_i \cos(2\pi f_i t) \quad 0 \leq t \leq T \quad (27)$$

is approximately equal to the weighted mean

$$\text{IWAIF}[s(t)] = \frac{\sum_i a_i^2 f_i}{\sum_i a_i^2} \quad (28)$$

The IWAIF model was applied to stimuli in some of the experiments for which the EWAIF model was able to predict the pitch differences. Table I shows the corresponding EWAIF and IWAIF values for a complementary Voelcker signal pair. The time-domain representation (11) was used to calculate the EWAIF and frequency domain representation (20) was used to calculate the IWAIF values. As noted in Feth et al. (1982), IWAIF differences are slightly smaller than EWAIF differences.

Iwamiya et al. (Iwamiya et al., 1983) studied the location of the principal pitch of FM-AM tones. These vibrato tones are generated by modulating a carrier both in frequency and amplitude with the same modulating signal. The stimuli used in these particular set of experiments consisted of a sinusoidal carrier at frequencies 440, 880, and 1500 Hz modulated by a triangular wave of frequency 6 Hz. Thus, if D_{AM} is the "degree of AM" and E_{FM} is the "extent of FM", the modulated signal for a carrier

frequency f_c is given by:

$$s(t) = [1 + D_{AM} m(t)] \cos[2\pi f_c t + 0.5 E_{FM} \int_0^t m(\tau) d\tau] \quad (29)$$

Listeners were asked to match the pitch of these modulated tones to a pure tone. The experiment was conducted for two cases. In the first case the degree of AM was set at unity and the extent of FM had values 0, 25, 50 and 100 cents. In the second case the extent of FM was a constant at 100 cents and the degree of AM had values 0.00, 0.50, 0.75 and 1.00. For each case, two sets of data were collected with the frequency and amplitude modulations in-phase and anti-phase respectively. EWAIF and IWAIF values were calculated for the case when the degree of AM is a constant. The results are shown in Fig. 1 for two carrier frequencies, 440 and 880 Hz. The curves with a positive slope represent FM and AM in-phase while those with a negative slope have their frequency and amplitude modulations 180° out of phase. IWAIF values are plotted along the dashed line while the dotted line corresponds to the EWAIF values. The regression equations for the localized principal pitch calculated by the authors is also shown as the solid line. It is clear from the figure that the weighted averages are reasonably close to the principal pitch values. Here, IWAIF differences are slightly larger than EWAIF differences.

Similar calculations were carried out for profile signals (Green, 1988; Green et al.,

1984). The standard consists of logarithmically spaced components drawn from a set of eleven frequencies ranging from 200 to 5000 Hz. Equal number of components are present on either side of the signal which is always at 1000 Hz. Thus, the three-component signal would have components ranging from 724 to 1380 Hz while the eleven-component complex would span the whole range from 200 to 5000 Hz. The results are shown in Fig. 2. The average threshold for a signal added to the middle component(1000 Hz) of the complex is plotted as a function of the number of components in the complex. The EWAIF and IWAIF values are superimposed on this graph with the solid line denoting differences in IWAIF values and the broken line denoting differences in EWAIF values as a percentage. Again, the difference between EWAIF and IWAIF values is very small.

IV. CONCLUSIONS

The intensity-weighted average of instantaneous frequency (IWAIF) model has been presented as an alternative to the envelope-weighted average of instantaneous frequency (EWAIF) model. Calculation of the EWAIF of a signal involves determining the envelope and the instantaneous frequency functions of the signal separately. This can be computationally cumbersome especially as the bandwidth of the signal gets wider. The IWAIF of a signal, on the other hand, can be expressed solely in terms of the magnitude spectrum of the signal. Such a frequency domain representation

provides a fast and efficient method to compute the IWAIF of a signal using the FFT algorithm.

The IWAIF model was tested on three sets of stimuli viz. Voelcker's complementary two-tone complexes used in experiments by Feth and co-workers (Feth, 1974; Feth et al., 1982; Feth and O'Malley, 1977), FM-AM tones used by Iwamiya et al. (Iwamiya et al., 1983), and profile signals used by Green et al. (Green et al., 1984). The performance of the IWAIF model was found to be comparable to that of the EWAIF model.

V. ACKNOWLEDGMENTS

This work was supported in part by a grant from the Air Force Office of Scientific Research and by a grant from the Ohio Regents Research Challenge Award.

APPENDIX A: EWAIF

The frequency domain representation of the EWAIF can be derived as follows. The EWAIF of $s(t)$ can be written as

$$\text{EWAIF} = \frac{1}{2\pi} \frac{\int_0^T |m(t)| \operatorname{Im} \left[\frac{m'(t)}{m(t)} \right] dt}{\int_0^T |m(t)| dt} \quad (\text{A1})$$

Consider the numerator,

$$\int_0^T e(t) f(t) dt = \frac{1}{2\pi} \int_0^T |m(t)| \operatorname{Im} \left[\frac{m'(t)}{m(t)} \right] dt \quad (\text{A2})$$

$$= \frac{1}{2\pi} \operatorname{Im} \int_0^T \sqrt{m(t) m^*(t)} \frac{m'(t)}{m(t)} dt \quad (\text{A3})$$

where $m^*(t)$ denotes the complex conjugate of $m(t)$

$$= \frac{1}{2\pi} \operatorname{Im} \int_0^T \left[\sqrt{m(t)} \right]^* \frac{m'(t)}{\sqrt{m(t)}} dt \quad (\text{A4})$$

$$= \frac{1}{\pi} \operatorname{Im} \int_0^T \left[\sqrt{m(t)} \right]' \left[\sqrt{m(t)} \right]^* dt \quad (\text{A5})$$

Applying the theorem for the Fourier transform of the derivative of a signal and invoking Parseval's theorem, the numerator can be expressed as

$$\int_0^T e(t) f(t) dt = \frac{1}{\pi} \operatorname{Im} \int_{-\infty}^{\infty} j2\pi f M_S(f) M_S^*(f) df \quad (\text{A6})$$

$$= 2 \int_{-\infty}^{\infty} f |M_S(f)|^2 df \quad (\text{A7})$$

where $M_S(f) = \mathcal{F} [\sqrt{m(t)}]$, is the Fourier transform.

Similarly, the denominator can be expressed as

$$\int_0^T |m(t)| dt = \int_0^T \sqrt{m(t)} [\sqrt{m(t)}]^* dt \quad (\text{A8})$$

$$= \int_{-\infty}^{\infty} |M_S(f)|^2 df \quad (\text{A9})$$

Hence

$$\text{EWAIF} = 2 \frac{\int_{-\infty}^{\infty} f |M_S(f)|^2 df}{\int_{-\infty}^{\infty} |M_S(f)|^2 df} \quad (\text{A10})$$

APPENDIX B: IWAIF

IWAIF can be variously expressed in terms of $s(t)$ and $\hat{s}(t)$. Using (17) the IWAIF of $s(t)$ can be written as

$$\text{IWAIF} = \frac{\int_0^T [s(t) \hat{s}'(t) - s'(t) \hat{s}(t)] dt}{\int_0^T s^2(t) + \hat{s}^2(t) dt} \quad (\text{B1})$$

$$= \frac{\int_0^T s(t) \hat{s}'(t) dt}{\int_0^T s^2(t) dt} \quad (\text{B2})$$

$$= - \frac{\int_0^T s'(t) \hat{s}(t) dt}{\int_0^T s^2(t) dt} \quad (\text{B3})$$

The last two relations were obtained by noting that the integral of the two terms in (B1) are equal.

In order to derive the frequency domain representation of IWAIF consider

$$\text{IWAIF} = \frac{\int_0^T |m(t)|^2 \operatorname{Im} \left[\frac{m'(t)}{m(t)} \right] dt}{\int_0^T |m(t)|^2 dt} \quad (\text{B4})$$

In the frequency domain the numerator can be expressed as

$$\int_0^T \epsilon(t) f(t) dt = \frac{1}{2\pi} \operatorname{Im} \int_0^T m(t) m^*(t) \left[\frac{m'(t)}{m(t)} \right] dt \quad (\text{B5})$$

$$= \frac{1}{2\pi} \operatorname{Im} \int_0^T m'(t) m^*(t) dt \quad (\text{B6})$$

$$= \frac{1}{2\pi} \operatorname{Im} \int_{-\infty}^{\infty} j2\pi f M(f) M^*(f) df \quad (\text{B7})$$

$$= \int_{-\infty}^{\infty} f |M(f)|^2 df \quad (\text{B8})$$

The expression for the Fourier transform of the differential of a signal as well as Parseval's relation were made use of in the foregoing simplification. Again, by Parseval's relation, the denominator is

$$\int_0^T \epsilon^2(t) dt = \int_0^T |m(t)|^2 dt \quad (\text{B9})$$

$$= \int_{-\infty}^{\infty} |M(f)|^2 df \quad (\text{B10})$$

Hence

$$\text{IWAIF} = \frac{\int_{-\infty}^{\infty} f |M(f)|^2 df}{\int_{-\infty}^{\infty} |M(f)|^2 df} \quad (\text{B11})$$

FOOTNOTES

¹Re denotes the real part operator

REFERENCES

- Anantharaman, J. N. (1992). A multichannel signal processing model for complex sound discrimination. Master's thesis, The Ohio State University, Columbus, OH.
- Anantharaman, J. N., Krishnamurthy, A. K., and Feth, L. L. (1991). Auditory processing of complex signals using the multichannel EWAIF. *Journal of the Acoustical Society of America*, 89(4):1938-1939. (abstract).
- Feth, L. L. (1974). Frequency discrimination of complex periodic tones. *Perception and Psychophysics*, 15(2):375-378.
- Feth, L. L. and O'Malley, H. (1977). Two-tone auditory spectral resolution. *Journal of the Acoustical Society of America*, 62(4):940-947.
- Feth, L. L., O'Malley, H., and Ramsey Jr., J. (1982). Pitch of unresolved, two-component complex tones. *Journal of the Acoustical Society of America*, 72(5):1403-1412.
- Feth, L. L. and Stover, L. J. (1987). Demodulation processes in auditory perception. In Yost, W. A. and Watson, C. S., editors, *Auditory Processing of Complex Sounds*, pages 76-86. Lawrence Erlbaum Associates, Hillsdale, NJ.

Green, D. M. (1988). *Profile Analysis: Auditory Intensity Discrimination*. Oxford University Press, New York, NY.

Green, D. M., Kidd Jr., G., and Mason, C. R. (1984). Profile analysis: Critical bands and duration. *Journal of the Acoustical Society of America*, 75:1163-1167.

Iwamiya, S., Nishikawa, S., and Kitamura, O. (1983). Perceived principal pitch of fm-am tones when the phase difference between frequency modulation and amplitude modulation is in-phase and anti-phase. *Journal of the Acoustical Society of Japan*, 5(2):59-69.

McGille, C. D. (1979). Hilbert Transforms and Analytic Signals. Unpublished notes on signal processing, Purdue University.

Voelcker, H. B. (1966a). Toward a unified theory of modulation-Part I: Phase-envelope relationships. *Proceedings of the IEEE*, 54(3):340-353.

Voelcker, H. B. (1966b). Toward a unified theory of modulation-Part II: Zero manipulation. *Proceedings of the IEEE*, 54(5):735-755.

von Helmholtz, H. L. F. (1954). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, page 165 and Appendix XIV. Dover Publications, New York, 2nd English edition.

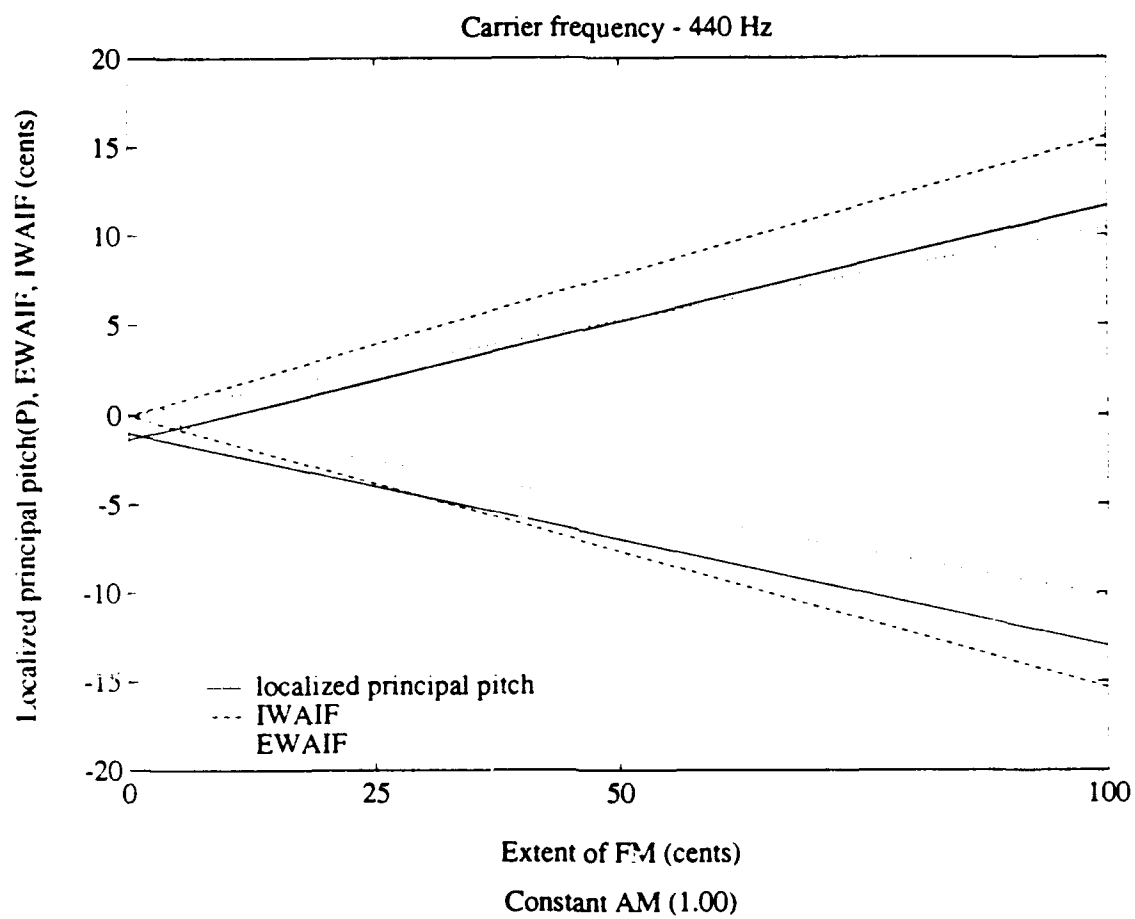
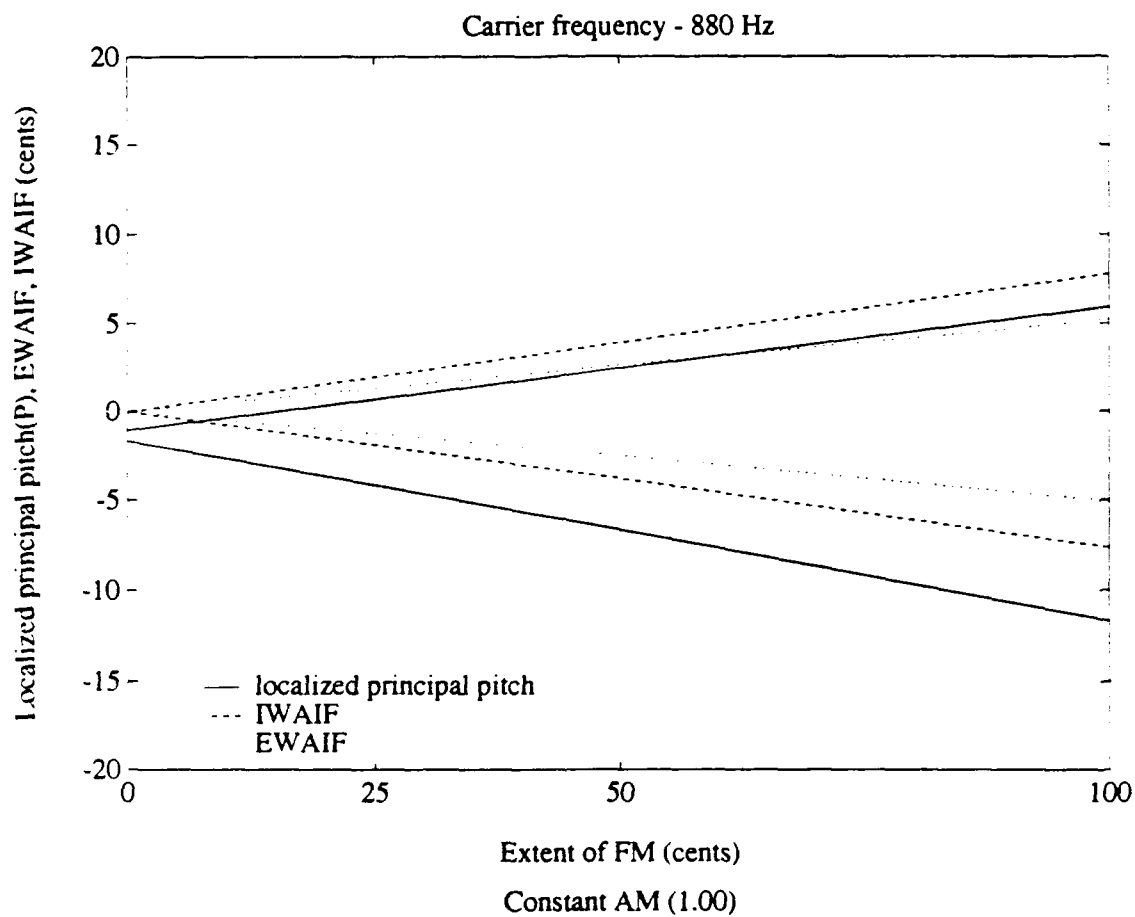


FIG. 1(a)



(b)

FIG. 1: Localized principal pitch, EWAIF and IWAIF of FM-AM tones as a function of the extent of FM with the frequency and amplitude modulations both in-phase (lines with positive slope) and anti-phase (lines with negative slope) (Iwamiya et al., 1983). (a) 440 Hz carrier frequency (b) 880 Hz carrier frequency

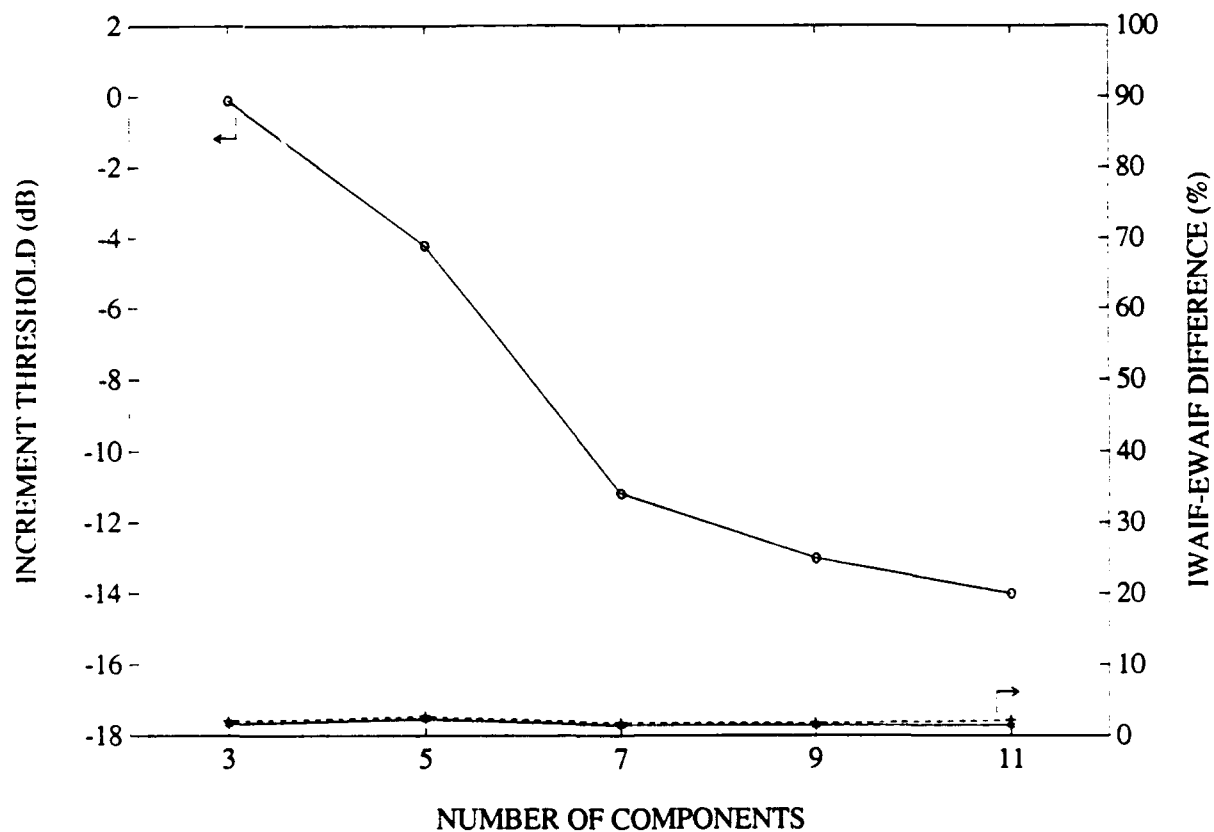


FIG. 2: EWAIF and IWAIF values calculated for profile signals (Green et al., 1984). EWAIF values are plotted as a dashed line ('+' symbols) while IWAIF values are plotted as a solid line ('*' symbols). The increment threshold for multi-component profile signals is also shown as a solid line ('o' symbols).

TABLE I: EWAIF and IWAIF values for complementary Voelcker signal pairs.

Signal	EWAIF	IWAIF
(1000 Hz, 71 dB) (1020 Hz, 70 dB)	1007.59	1008.82
(1000 Hz, 70 dB) (1020 Hz, 71 dB)	1012.41	1011.18

Using the Ariel DSP-16 as a Signal Generator for
Psychoacoustics Experiments

Chien-Yeh Hsu and Lawrence L. Feth

Division of Speech and Hearing Science

The Ohio State University

110 Pressey Hall, 1070 Carmack Road

Columbus, Ohio 43210-1002

Phone Number: 614-292-8207

E_mail:hsuc@shs.ohio-state.edu

done 16

ABSTRACT

A set of TMS320 assembly programs have been implemented to generate different stimuli used in psychoacoustic experiments. **GENTONE.ASM** generates single frequency pure tones, step tones and glide signals. **MULTI.ASM** can be used to generate multi-component signals. Finally, **CLICK6.ASM** can generate two click sequences with a specified delay time between them. By using these TMS320 programs, stimuli can be generated in real time (in **GENTONE.ASM** and **CLICK6.ASM**) or much faster than the conventional microcomputers (in **MULTI.ASM**). The TMS320 programs can be invoked by several high level programming languages including C, PASCAL, FORTRAN and BASIC. All these programs have been tested on the *Ariel DSP-16* board installed in the IBM PC compatible computers.

Introduction

Generating accurate and stable stimuli is usually important in many hearing experiments. Since microcomputers have been widely used to control and monitor laboratory events, it is very convenient to generate or digitize analog signals by using a digital-to-analog converter (DAC) and analog-to-digital converter (ADC).

Because multi-listener adaptive paradigms and/or roving frequency paradigms are included in many hearing experiments, it is often necessary to generate "real time" stimuli rather than stored signals. Using stored signals will require disk I/O time. This longer time delay often can not satisfy the requirements of many experiments which use multiple listener, adaptive-tracking, roving-frequency procedures. In addition, stored signals usually consume lots of disk spaces.

This report describes a set of programs which can generate signals in real-time for use with IBM compatible microcomputers. The programs discussed in this report are implemented in TMS320 assembly language (Texas Instruments Incorporated, 1987) and can run independently on the Ariel DSP-16 board (Ariel corporation, 1987). The TMS320 assembly programs are very fast and can communicate with many high level languages. Therefore, they provide a powerful utility for generating signals in

"real time". In this report, we will introduce several TMS320 programs which generate pure tones, step tones, glides, multi-component signals, and clicks.

The Ariel DSP-16 provides a very good hardware and software environment for developing our application programs. The resident monitor (RESMON) is a very useful tool for writing a TMS320 application program. RESMON is used to ease the task of constructing all the programs described in this article. It is always assembled to fill 1024 words (from the address 0 to 1023). For more information about the DSP-16 and RESMON, please refer to the operating manual (Ariel Corporation, 1987).

All of the programs described in this report have been used with TURBO PASCAL (Borland International, 1989) in our laboratory to control experiments such as "frequency discrimination" and "complex sound discrimination".

Pure tone, step tone and glide signals

To generate single frequency tones, step tones or glides, we use the *Table Lookup* method. A one-cycle sine wave (with frequency $f = 1$, at a sample rate of $R = 50000$ samples per second) is saved in the data buffer having $L = 50,000$ memory locations. In order to convert digital samples, the TMS320 program has an address variable that points to the particular memory location that is converted to analog form at a given instant. After each

sample conversion, the address variable is incremented by Δ to point to the next sample in the data buffer. When the address variable (*Addr*) is larger than 49,999, the address variable is reset to the offset value (*Addr* - 50000). Therefore, a waveform is generated by the continuous recycling of the sampled sine data. If we set the sample rate of the DSP-16 to 50 KHz, the output frequency f' is equal to Δ . The frequency resolution is, therefore, 1 Hz.

The TMS32025 program used for generating different tones is called **GENTONE.ASM**. This program can generate steady tones, step tones and glides. If we keep the increment value Δ fixed, the output will be a single frequency signal (steady tone). However, if we increase or decrease Δ linearly (the increment or decrement of Δ is equal to one for each step), the output will be a glide signal with rising or descending frequency. For example, if Δ changes smoothly from 400 to 600, a glide tone starting from 400 Hz, ending at 600 Hz is generated. In addition, if the increment of Δ is a fixed number rather than one for each step, a step tone is generated.

Insert Figure 1 about here

Figure 1 shows an example of these three types of tones (duration = 120 ms): a 400 Hz steady tone, a glide

from 900 Hz to 1140 Hz having a *slope of 2 Hz per milli-second*, and a step tone from 600 to 840 Hz having seven steps and the same slope as that of the glide. Note that the *step length* (in ms) of the first and last steps (short steps) is the half of the *step length* of the "middle" steps (long steps). This definition is necessary, so that the slope (see the dot line in Figure 1) can be represented correctly. Also, the glide tone is a special case of step tones with (*frequency excursion + 1*) steps and a *step height* of 1 Hz.

Insert Figure 2 about here

Often, the output of the DSP-16 is passed through an analog gate. Without the gating, the waveform will produce an abrupt amplitude transition at the signal onset and offset. These are often heard as clicks. A software gating mechanism is incorporated into **GENTONE.ASM** so that external gating is not necessary. The software gate is defined by a gating envelope (see Figure 2). The output waveform is then multiplied by the gating envelope. Figure 2 shows that a sine wave is gated by a cosine squared gating envelope. A parameter in the program can be used to determine if the hardware gates are available or the software gate will be used. By using the software gate, the program can generate signals

having smooth envelopes. However, the gating envelope must be loaded into the data buffer before the signals can be generated.

The first part of **GENTONE.ASM** is the RESMON monitor supplied by Ariel company and the second part is the user function, which starts at the address 1024. One user function (function #0) is defined in **GENTONE.ASM**. In order to use the program, the output sampling rate of the DSP-16 should be set to 50 KHz. When the high level program calls the user function to generate signals, ten parameters have to be sent. The parameters are described in Table 1.

Insert Table 1 about here

Note that when the sample rate is 50 KHz, the parameter *step length* defined in Table 1 can be written as

$$\text{Step length} = \begin{cases} \frac{50 \times \text{Duration}}{\text{step number} - 1} & \dots \text{long step} \\ \frac{50 \times \text{Duration}}{2(\text{step number} - 1)} & \dots \text{short step} \end{cases} \quad (1)$$

where *Duration* is the signal duration in milli-second. Furthermore, step length usually does not need 32 bits (16 bits are enough for 65,535 samples). However, we use 32 bits to represent step length for the case of very long steps and to give more flexibility to the users.

Before the main program can call **GENTONE.ASM**, a sine table has to be loaded into the DSP-16 data buffer. The sine table should be a one-cycle sine wave containing 50,000 samples, in locations from 0 to 49,999 in the data buffer. The DSP-16 has at least a 256K data buffer, so the sine table will occupy one-fifth of that buffer. Also, one can call the RESMON command CMD11 - "Write to buffer RAM" to load the sine table. A TURBO PASCAL example used for loading the sine table will be given latter. If the software gate is used, the gating envelope should be loaded into the data buffer. The address of the gating envelope should be at 50,000. The length of the gating envelope is the same as that of the signal. Note that the "real" length of the envelope should be computed by using the *long-step-length* and the *short-step-length*

$$\text{envelope length} = (\text{long step length}) \times (\text{step number} - 2) + (\text{short step length}) \times 2 \quad (2)$$

In fact, by using the gating envelope, we can modify the output waveform to any shape of envelope.

After loading the sine table and the gating envelope (if it is necessary), one can issue the RESMON command (CMD12) to invoke the user function. Since 10 parameters will be passed to the function, the code format here for CMD12 is B280h. Finally, the parameters described in Table 1 are sent to the DSP-16, one by one. When the function is finished it sends an FFFFh value back to the

PC. This value must be read from the host port into the PC before any further host port communication can be executed.

GENTONE.ASM has been included in many TURBO PASCAL programs, which control the experiments designed for the study of frequency discrimination in our lab (Neill, 1990). Because the stimuli are generated in real time, we can include adaptive procedures and roving level methods in the experiments. Also, multi-subject adaptive procedures are possible because of the high speed of the TMS programs. This can save a lot experiment time and no disk space is necessary to store the stimuli.

Insert Figure 3 about here

An example of a TURBO PASCAL program segment is shown in Figure 3. Here we assume the TMS program **GENTONE.HEX** (machine code of **GENTONE.ASM**) has been loaded into the DSP-16's program/data memory and all the variables are declared correctly. The example consists of two parts, the first part shows how to load the sine table into the DSP-16's data buffer RAM, and the second part is the program for calling the "GENTONE function". In the first part of this example, the sine table is a binary file of integers in IBM PC format. Also, the output sampling rate has to be initialized to 50 KHz.

Multi-component complex tones

A complex periodic tone is a signal consisting of multiple components each with different frequencies and amplitudes. A *multi-component signal* can be written as

$$s(t) = \sum_{i=0}^m A b^i \cos(2\pi(F_a + i f_b) t). \quad (3)$$

The line spectrum for this example is shown in Figure 4.

Insert Figure 4 about here

The use of the TMS320 to generate multi-component complexes is not as simple as that described for generating pure tone and glide signals. A set of TMS320 functions has been developed for the purpose of generating multi-component signals. This set of functions is described in Table 2. Note that the function "FIND_PEAK" and "NORMALIZE" are modified from the program (S1 Data Acquisition Application) incorporated with the DSP-16, provided by Ariel company (Ariel corporation, 1987).

Insert Table 2 about here

All the functions, except "FIND_PEAK", will return a number valued FFFFh to the PC when they are finished. The

returned number must be read out by the host computer before any further host port communication can be issued.

After we specify the functions in Table 2, the algorithm for producing multi-component signals can be presented below:

1. Create a sine table in the DSP-16's data buffer RAM from location 0 to 49999, and a copy of the sine table at addresses from 50000 to 99999. This is necessary because of the design of the TMS320 functions described above.
2. Clear the signal buffer. The signal buffer starts at the address of 100000. The length of the signal buffer depends on the duration of the stimulus and the signal buffer will end at the address of $100000 + \text{length} - 1$.
3. Generate one component of the signal according to the frequency of that component using "GEN_COMPONENT".
4. Modify the amplitude of the component. This can be done by multiplying values in the signal buffer by a constant using the "NORMALIZE" function.
5. Add the component generated in step 3 and 4 to the signal buffer.
6. Repeat step 3 to 5 until all the components of the signal have been generated.

The above algorithm requires sampling, normalizing and adding each of the components in turn, therefore, the signal can not be generated in real time, despite the TMS32025 processor's speed. This TMS320 program is still

much faster than any high level language for generating multi-tone complexes. For example, it takes 1.71 seconds to generate a 125 ms ten-component signal using the TMS320 program. But, it takes 6.21 seconds to generate the same signal using TURBO PASCAL on our 386 machine with Cyrix CX-83D87 math processor (Cyrix Corporation, 1990). If there were no math co-processor, the time needed to generate the same signal would extend to 33.23 seconds. For this comparison, we used the same algorithm in the TMS320 program and the TURBO PASCAL program. The time consumption of disk input/output and loading the sine table are not included.

All of the functions discussed in Table 2 have been incorporated into a TMS32025 program called **MULTI.ASM**. The RESMON monitor occupies the first 1K byte as usual and is followed by seven user functions. **MULTI.ASM** has been used in our lab to generate multi-component signals for "Profile Analysis" (Whitelaw et al., 1991) and "Complex Sound Discrimination" experiments.

Click signals

It is very convenient to use the **DSP-16** to generate pulse-like signal. A pulse can be generated by sending a sequence of number, which contains only 0 and 32767 (the maximum 16-bit positive number), to the DAC.

A TMS32025 program, **CLICK6.ASM**, has been developed in our lab to generate dual-channel click sequences. For

the requirement of our experiment, the signal output of the program contains a dual-channel click train followed by a silent delay and a test click signal. A typical signal output and its parameters are shown in Figure 5. This signal is very similar to the stimuli used in Freyman's study of the precedence effect (Freyman et al., 1991).

Insert Figure 5 about here

Like the other programs, **CLICK6.ASM** consists of two parts. The first part is the RESMON monitor, which occupies the first 1K byte of the program memory, and the second part is the user functions. In **CLICK6.ASM**, only one user function is declared: the user function #0, which will generate the signal output. User function #0 can be called by several high level programming language. Ten parameters should be sent to the function. The procedure for calling the function is described as follows.

First, call the RESMON command (CMD12) to activate the user function. Because ten parameters need to be sent to DSP-16, the code format issued for CMD12 is B280h. Following CMD12, ten parameters are sent to DSP-16. These parameters and their definitions are shown in Table 3.

Insert Table 3 about here

Note that the time measurement in the table is equal to 50 times the duration in milli-second. This is because the sampling rate is set to 50 KHz in **CLICK6.ASM**.

A TURBO PASCAL example for calling this user function is given in Figure 6. In this example, we assume that all the variables and procedures have already been declared.

Insert Figure 6 about here

Several TURBO PASCAL programs using **CLICK6.ASM** have been implemented in our lab to control the experiments in the study of "precedence effect" (Wallach et al., 1949). There are several advantages of using **CLICK6.ASM** instead of using pre-recorded stimuli. First, the user can control all the parameters of the signal very easily in high level programming language to generate different stimuli. Second, since the parameters can be modified in the program, it is possible to design adaptive procedures and roving level techniques in experiments. Third, the silent delay is controlled by the TMS32025 processor, so we can specify a very precise duration for the silent delay. For example, if the sampling rate is 50 KHz, the

duration can be precise to 0.02 milli-second (the reciprocal of the sampling frequency). Finally, no disk space is necessary for saving different stimuli, because the stimuli are generated in real time by DSP-16. All of the programs have been tested on IBM PC/AT and 80386 compatible machines.

System requirements

All of the TMS320 programs described in this report are designed for use on the Ariel DSP-16 Data Acquisition system (Ariel corporation, 1987) equipped with at least 256K words of data buffer RAM. The first 1K bytes of the programs is always the RESMON monitor provided by the Ariel company. The user functions are assembled starting at the address 1024. All the TURBO PASCAL examples shown in this report are tested on IBM PC/AT and 386 compatible machines. The TMS320 programs can also interface to some other high level languages such as *Microsoft C* and *BASIC* (Ariel corporation, 1987).

Availability of programs

The following listing are available from the authors upon request: the source code and assembled machine code of the TMS320 programs described in this report, TURBO PASCAL program example of initializing DSP-16 and loading programs to DSP-16, and detailed TURBO PASCAL examples of using these TMS320 programs. The first author may be contacted via E-mail at hsuc@shs.ohio-state.edu.

Acknowledgment

This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR089-0227.

Table 1

Parameters Used for User Function #0 in GENTONE.ASM

Parameter	Description
Start frequency (Hz)	The start frequency of the signal.
Step height (Hz)	The frequency distance between two steps of the signal. Note that if a steady tone is wanted, step height is equal to zero. The step height is equal to one for glide tones.
Step number	The number of steps in a signal. Note that a steady tone has a step number of one and the step number of a glide tone is equal to the <i>frequency excursion</i> plus one.
Long step length (high)	The high 16 bits of the <u>number of samples</u> of the long steps. The middle steps of a signal are assigned to <i>long steps</i> (see Figure 1).

(table continues)

Parameter	Description
Long step length (low)	The low 16 bits of the <u>number of samples</u> of the long steps.
Short step length (high)	The high 16 bits of the <u>number of samples</u> of the short steps. The <i>short steps</i> are the first and last steps of a signal.
Short step length (low)	The low 16 bits of the <u>number of samples</u> of the short steps.
Rising or descending	This value is equal to 0 for rising tones and 1 for descending tones. This value has no effect on a steady tone.
Output channel	Define the output DAC to be channel A, B or both channels. Send 1 for channel A only, send 2 for channel B only, send 3 for channel A and B (diotic).
Software gate	Reset this value to 0 to turn on the software gating operation. Set this value to 1 if no software gate is used.

Table 2

Functions used for generating multi-component complexes.
Note that all the addresses are 32-bit long-integer
consisting of high 16 bits and low 16 bits. Thus an
address variable is defined by two 16 bit words.

Function	Description
DAC	Digital to analog conversion function. Output signals to both channels. Code format = B200h. Eight parameters : start address and stop address of channel #1, and start address and stop address of channel #2. Return an FFFFh value to PC when function is done.
FIND_PEAK	Find the peak value for a data segment. Code format = B101h. Four parameters : Start address and stop address of the data. Return the peak value and a 32-bit peak address (low and high). Note that this function must be executed before calling NORMALIZE.

(table continues)

Function	Description
NORMALIZE	<p>Normalize signals in the data buffer to a given maximum value. Code format = B342h. Twelve parameters : start and stop addresses of channel #1, start and stop addresses of channel #2, channel flag (1 for channel #1, 2 for channel #2, 3 for both channels), integral scale factor for channel #1, integral scale factor for channel #2, fractional multiplier for channel #1, and fractional multiplier for channel #2. Return an FFFFh value when function is done. The fractional multiplier is defined as the fractional part of scale factor times 32767.</p>

(table continues)

Function	Description
CLEAR_BUFFER	Reset the signal buffer to zero. Code format = B103h. Four parameters : start and stop addresses of the buffer. Return an FFFFh value when function is done.
GEN_COMPONENT	Generate a single frequency component. Code format = B144h. Five parameters : start and stop addresses of the signal buffer and the frequency of the component. The method used to generate signals in this function is very similar to that used in generating pure tone, which is described in the previous section. Return an FFFFh value when function is done.

(table continues)

Function	Description
COPY_DATA	Copy a block of data to another block. Code format = B245h. Nine parameters : the start and stop addresses of the original, the start and stop addresses of the destination, and a flag. If the flag is 1, opposite data will be copied. If the flag is 0, same data will be copied. Return an FFFFh value when function is done.
ADD_BLOCK	Add two blocks of data. This function can be written as $signal1 = signal1 \pm signal2$. Code format = B246h. Nine parameters : the start and stop addresses of <i>signal1</i> , the start and stop addresses of <i>signal2</i> , and a flag. The flag is equal to 0 for plus, 1 for minus. Return an FFFFh value when function is done.

Table 3

Parameters used for user function #0 in CLICK6.ASM

Parameter	Definition
pattern	Define the pattern for one cycle of the two-channel pulse sequences. 1 for high output, 0 for low output. For the signal described in figure 4, the value of pattern is 33, which is 00100001 in binary.
repeat time	Define the number of cycles of the condition pulse, when there is a condition pulse. This value is 1, when there is no condition pulse.
segment	The number of segments in one cycle of the pulse sequence. For the signal in figure 4, the value of segment is 4.
lag1 (50*ms)	Delay of the second pulse of the test signal. Note that the value sent to the DSP-16 is (lag1 - pulse_width).

(table continues)

Parameter	Definition
ici (50*ms)	Inter click interval is the period of the pulse train. Note that the value sent to the DSP-16 is (ici - lag1 - pulse_width).
pulse_width (50*ms)	The width of the click.
p1	Send 1 if B channel is active for test signal. Send 0 if B channel is inactive for test signal.
silent_delay (50*ms)	Delay between the condition pulse train and the test signal. Note that the value sent to DSP-16 is (silent_delay - ici)/2, because the TMS32025 program is designed in this way.
p2	Send 1 if condition pulse train is activated. Send 0 if there is no condition pulse train.

(table continues)

Parameter	Definition
lag_cond (50*ms)	Delay of the second pulse train of the condition signal. This parameter is defined as the same way of lag1. The value sent to the DSP-16 is (lag_cond - pulse_width).

Figure Caption

Figure 1. An example of the three types of tones generated by GENTONE.ASM. This figure shows a 400 Hz steady tone, a 7-step step tone having a step height of 40 Hz, and a glide tone from 900 to 1140 Hz.

Figure 2. (a)A sine waveform, (b)A cosine squared gating envelope, (c)The modified sine waveform.

Figure 3. A segment of TURBO PASCAL program including examples of loading sine table into the data buffer and calling GENTONE.ASM.

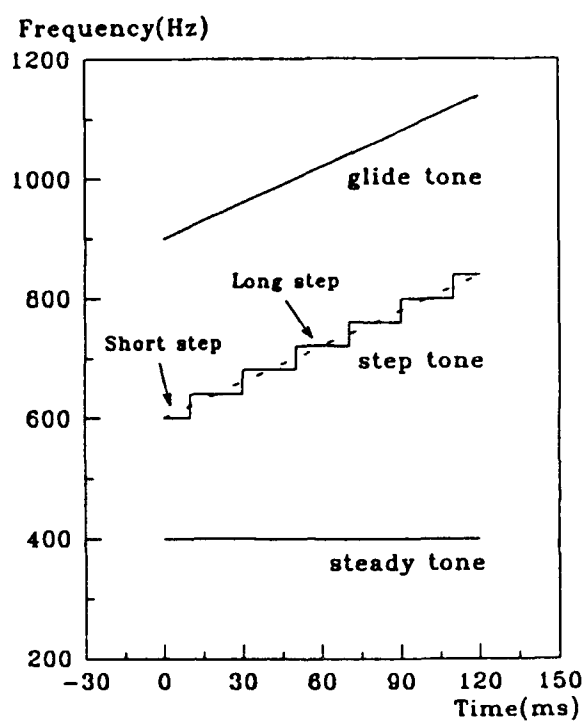
Figure 4. Line spectrum for a multi-component signal.

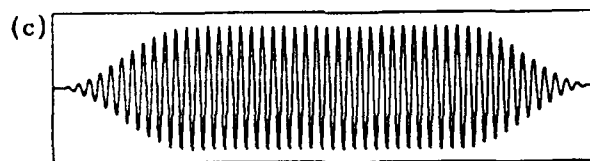
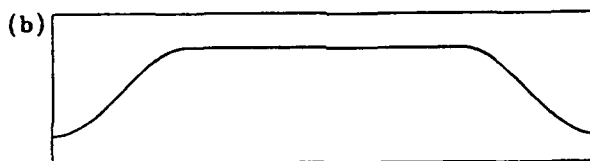
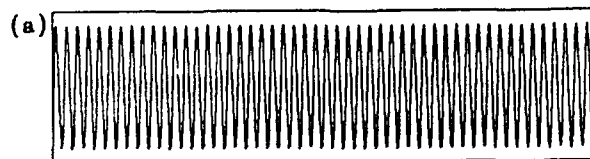
Figure 5. A typical signal generated by CLICK6.ASM. The parameters associated with the signal are also shown.

Figure 6. A TURBO PASCAL example for calling CLICK6.ASM.

References

- Ariel Corporation. (1987). *Operating Manual for the DSP-16 Data Acquisition Processor* (4th ed.). New York, New York: Author.
- Borland International. (1989). *Turbo Pascal Reference Guide*. Scotts Valley, CA 95066-0001: Author.
- Cyrix Corporation. (April, 1990). *FasMath™ 83D87 User's Manual*. Richardson, TX 75085-0118: Author.
- Freyman, R. L., Clifton, R. K. and Litovsky R. Y. (1991). "Dynamic processes in the precedence effect," *J. Acoust. Soc. Am.* 90, 874-884.
- Neill, M. E. (1990). Influence of Randomly Chosen Starting Frequency on Frequency-Glide Direction Discrimination. Master's thesis, The Ohio State University, Columbus, OH.
- Texas Instruments Incorporated. (1987). *Second-Generation TMS320 User's Guide*. Houston, Texas 77001: Author.
- Wallach, H., Newman, E. B. and Rosenzweig, M. R. (1949). "The precedence effect in sound localization," *Am. J. Psychol.* 62, 315-336.
- Whitelaw, G. M., Hsu, C. Y. and Feth, L. L. (1991). "Dichotic profile analysis," *J. Acoust. Soc. Am.* 89(4), 4PP3.





Time

DICHOTIC AUDITORY PROFILE ANALYSIS

Gail M. Whitelaw*

Lawrence L. Feth

Division of Speech and Hearing Science; The Ohio
State University; Columbus, Ohio 43210

*currently affiliated with Department of Speech
and Hearing Sciences; Indiana University

Received:

To date, a limited number of studies have used a dichotic stimulus configuration presentation in profile analysis experiments. This study used a dichotic stimulus configuration for profile analysis which had not been previously used. Profile stimuli were presented both diotically and dichotically. The diotic signal consisted of twenty-one components with equal logarithmic spacing presented to both ears. The profile signal was created by incrementing the odd numbered components of the signal by n dB. The dichotic signal was created by presenting the odd numbered components to the right ear and the even numbered components to the left ear. Two experiments were performed. The first experiment was to determine if profile analysis abilities differed when stimuli were presented dichotically. The second experiment was performed to determine if subjects might be using interaural intensity difference cues instead of a profile analysis strategy during the dichotic listening task. Results suggested that most subjects were able to perform the dichotic profile analysis task, with significant differences noted between diotic and dichotic results. Results of the second experiment suggest that profile analysis and interaural intensity differences do not appear to result from the same mechanism.

PACS Numbers: 43.66E; 43.66R

INTRODUCTION

Studies of auditory profile analysis have demonstrated that listeners are able to discriminate very small intensity increments added to a multitonal background complex, supporting the concept that auditory cues remote from a target signal may assist in discrimination of that signal. During the past decade, a number of stimulus characteristics have been investigated within the profile analysis paradigm. The majority of these studies have used either a monotic or diotic stimulus presentation. To date, very few studies have used dichotic stimulus presentation.

The limited number of studies that have used a dichotic stimulus have generally presented a single center component to one ear and twenty background components to the other ear simultaneously (e.g. Green & Kidd, 1983; Bernstein & Green, 1987) (Figure 1a). The results of these studies have suggested that listening performance was inferior for a dichotic listening condition when compared to a monotic listening condition. A conflicting result was presented by Fantini, Schooneveldt, & Moore (1989). When a dichotic profile configuration with the target component was presented to one ear and four background, or "flanking," components were presented to the other ear simultaneously (Figure 1b), they reported that the dichotic presentation of the background components did not significantly affect

profile analysis results when compared to a monotic presentation mode. Fantini et al. suggest that listeners were able to combine information across ears in order to develop a profile, at least for the conditions reported in their study.

A potential problem with the type of dichotic presentation methods previously used in profile analysis tasks is that listeners may be able to perform the task from the information presented to a single ear. The listener may be able to use information presented to each ear separately to make a correct response, without needing to combine the information between the two ears. In addition, this stimulus configuration may not provide a fused image that the listener can process binaurally.

A type of dichotic profile has been developed for this study in which the stimulus presented to either ear alone provides no profile information and the target signal can only be recognized when stimuli are combined between the two ears.

The development of this type of stimulus configuration is based on cyclopean perception experiments in vision proposed by Julesz (1971). In these experiments, different information is presented to each eye simultaneously, and the target will only be perceived if the inputs to the two eyes are combined; if the target is perceived when a stimulus is presented to either eye individually, Julesz suggests that the stimulus is not being perceived by a central mechanism.

A limited number of auditory analogues of cyclopean perception have been reported. Huggins and Cramer (1958) demonstrated that pitch can be perceived as a result of dichotic interaction of noise stimuli presented to both ears simultaneously. Houtsma and Goldstein (1972) illustrated that a mechanism exists for combining information from the two ears to form a "central spectrum". The findings of these studies suggest that a mechanism exists for combining information which is presented between the two ears. Experiments such as these, which simulate cyclopean aspects of the auditory system, may give insight into the central aspects of how information is extracted from complex auditory stimuli. Julesz (1971) suggests that despite the difficulties in developing techniques to investigate the "cyclopean cochlea", further definition of the location of a cyclopean ear is necessary. The use of dichotic profile stimuli is of interest in order to understand the central versus peripheral nature of the profile analysis task.

The present study was designed to use a cyclopean type of stimulus to explore the auditory system. A type of dichotic stimulus was used for the profile analysis task which had not been previously used in other dichotic profile analysis studies. In this study, the profile stimulus was split between the two ears in a manner in which the profile cannot be perceived by the listener when the information was presented to either ear alone (Figure 1c). The perception of

the profile occurred only when the stimuli were combined between the two ears. It was postulated that splitting the component frequencies between the two ears may allow the listener to develop a better fused image to be processed binaurally than in the stimuli previously used in profile analysis experiments using dichotic stimuli. This is supported by Dye (1990), who suggested that listeners perceived a single, fused lateralized image even when multiple components of complex signals, greater than one critical band apart, were presented dichotically. In addition, it was of interest to use a dichotic profile stimuli in order to further understand the possible central nature of the profile analysis task.

The major purpose of this study was to determine if profile analysis abilities differ when stimuli are presented dichotically.

EXPERIMENT 1: Dichotic Profile Analysis

I. METHODS

A. Subjects

Four young adult females, ranging in age from 19 to 32 years, participated in this study. All subjects had normal hearing acuity and normal middle ear functioning bilaterally. Three of the four listeners were college students paid for their participation in this study; the fourth listener was one of the authors. One of the subjects had prior experience as a listener in psychoacoustic tasks, and the other three

subjects participated in the pilot experiment for this study (Whitelaw, Hsu, and Feth, 1991). Each subject had at least twenty hours of practice in the task to assure that stable performance was achieved prior to initiation of data collection.

B. Stimuli

Signals were digitally generated on-line using an Ariel DSP-16 signal processing subsystem housed in a Zenith Z159 microcomputer. These signals were converted into analog waveforms for each listening interval using a two channel 16-bit digital-to-analog converter at an overall sampling rate of 50,000 Hz per channel (Hsu and Feth, 1992).

Profile stimuli were presented both diotically and dichotically. The diotic signal consisted of twenty-one components with equal logarithmic spacing (Figure 1c). The mean level of the signal was 50 dB SPL per component. Level per component was equal in the standard signal. To create the profile, the odd numbered components were incremented by n dB, where $n = .5, 1, 2, 3, 4, 6, \text{ or } 7$ dB. All twenty-one components were presented to both ears simultaneously. The dichotic signal was created by presenting the odd numbered components to the right ear and the even numbered components to the left ear. Stimuli were presented with a roving level to limit the listeners ability to base discrimination decisions on simple stimulus level changes (Berliner and Durlach, 1973). A roving

range of 20 dB (\pm 10 dB), varying in 1 dB steps was employed.

Three center frequencies--500 Hz, 1000 Hz, and 2000 Hz--were employed for both diotic and dichotic stimulus presentation. Equal logarithmic spacing, corresponding to a frequency ratio of 1.175, was used for the three center frequencies. Each signal was presented for 200 ms, with an inter-trial interval of approximately 500 ms.

Listeners were tested in separate single-walled sound-isolated rooms. A group of three listeners were able to listen at the same time. Each listener was seated facing a monitor and computer keyboard (Radio Shack Color Computer II).

Signals were presented binaurally via Sennheiser HD414SL headphones. Listening intervals were indicated with visual cues and visual feedback displaying the correct response was provided immediately following each trial. All subjects listening at the same time heard the same signals and their responses were recorded separately. Recording of responses and visual feedback was controlled by the Zenith Z-159 microcomputer.

Data collection was carried out in blocks of 50 trials, with listeners taking a break after every three blocks. The intensity increment step was held constant over each group of three blocks. Short breaks were provided when three blocks were completed. Longer breaks were given after six blocks

were completed. Data from approximately 900 trials, or 6 groups of three blocks, were obtained in each daily session.

A 2 cue-2 alternative forced-choice (2Q-2AFC) procedure was used for this study. For the profile analysis task, a standard profile, consisting of twenty-one components with equal level per component, was always presented in intervals one and four. The signal was generated by adding an increment to the level of the odd-numbered components. It was presented in either interval two or three with equal probability. A roving level paradigm was employed, with a roving range of 20 dB (± 10 dB) varying in 1 dB steps. In this paradigm, the overall intensity level for each trial was randomized, minimizing the listener's use of intensity anchors.

The listeners' task was to determine which of the two center intervals of the 2Q-2AFC paradigm contained the target signal. Listeners indicated their responses by pressing the appropriate key on the microcomputer keyboard.

To assure that listeners required stimulus input from both ears in order to discriminate a "profile", the eleven component signal with the intensity increment added was presented to the right ear only for a separate block of trials. A 6 dB increment was added, since in the dichotic case, a majority of the listeners could easily discriminate a 6 dB intensity increment. In this task, all subjects' performance fell to chance over at least 300 trials.

III. RESULTS

Percent correct discrimination was plotted as a function of intensity increment for each of the three center frequencies. Each percent correct discrimination score was the average of 300 trials (6 blocks of 50 trials). All six blocks fell within one standard error of the mean of the proportion for a binomial variable (Krajewski and Ritzman, 1990). This required that the percent correct score for all six blocks be within $\pm 8\%$ of the mean percent correct score for the six blocks. If all six blocks did not meet this criterion, another six blocks were collected and the process was repeated.

To facilitate fitting a line to the data, percent correct discrimination scores were converted to d' values for a 2AFC task (McNichol, 1972). d' values were plotted as a function of intensity increment for each center frequency. A simple regression line was fit to each psychometric function. Psychometric functions at each frequency for individual listeners are presented in Figures 2a-d. Discrimination thresholds were derived from the linear regression formula which was used to determine the intensity level corresponding to a d' of .95, or 75% correct value. Discrimination thresholds for individual subjects are presented in Table 1.

Several points are noted from inspection of the profile results for individual subjects. For the diotic listening conditions, all subjects were able to achieve threshold for

all frequency conditions. However, in the dichotic listening condition, only three of the four subjects achieved 75% performance. The fourth subject was unable to reach 75% performance on the dichotic condition for any of the center frequencies. This observation is supported by findings in previous experiments in profile analysis, in which considerable variability for individual performance on this task has been noted (Green, 1988b).

Visual inspection of the psychometric functions reveals that the dichotic results are shifted to the right relative to the diotic results for all frequency conditions and for all listeners. Diotic performance was compared to dichotic performance for each center frequency for the listeners who were able to perform both the diotic and dichotic task. This comparison was accomplished by using a t-test designed to determine the significance of differences between independent slopes (Cohen and Cohen, 1983). The results of this t-test demonstrated that the slopes of the lines were significantly different at 500 Hz and 1000 Hz for all three subjects. The slopes at 2000 Hz were significantly different for one of the subjects (S2), while a significant difference was not observed for the other two subjects (S1 and S3).

d' values were averaged across the three subjects who were able to achieve threshold for both the diotic and dichotic listening task. Although the results of the fourth

subject are not included in this analysis, they will be included in the general discussion of results.

Psychometric functions averaged across the subjects for each center frequency are shown in Figures 3a-c. Simple regression lines were fit to these data and are also shown in these figures. Diotic discrimination thresholds ranged from 2.0 to 2.2 dB and dichotic discrimination thresholds ranged from 4.5 to 4.8 dB across the test frequencies.

A three way analysis of variance (ANOVA) was used to determine effects of stimulus presentation (diotic vs. dichotic) and effects of center frequency. This statistical analysis confirms the conclusions obtained through visual inspection of the data. That is, significant differences were observed between the diotic and dichotic stimulus conditions ($F=9.21$, $p>.01$) and between listeners ($F=69.02$, $p>.001$). No significant differences were noted among the three frequency conditions. In addition, significant listener-by-stimulus or listener-by-frequency interactions were not found.

In discussing preliminary results of this study, Bernstein (1990) suggested that subjects might be using interaural intensity difference cues instead of a profile analysis strategy during the dichotic listening task. To test this premise, a second experiment was performed.

EXPERIMENT 2: INTERAURAL INTENSITY DIFFERENCES

The ability of the auditory system to detect very small differences in intensity for stimuli presented to the two ears

is well documented. Nuetzel (1982) reported that when interaural intensity difference (IID) threshold were measured for pure tones, listeners were sensitive to small differences in intensity even when frequencies of the two tones were separated by more than half an octave. He noted a frequency effect for IID performance, since IID thresholds were found to be lower for higher frequency stimuli. Based on Nuetzel's research, Bernstein (1990) has suggested that when dichotic profile analysis stimuli are presented to listeners, they may be able to use a binaural intensity summation cue and focus on an interaural intensity difference between the center frequency and adjacent component instead of performing an analysis of spectral shape.

I. METHODS

A. Subjects

The same four subjects that participated in Experiment I participated in this experiment.

B. Stimuli

Pure tone stimuli for binaural interaural intensity difference thresholds were digitally generated on-line, using the dichotic profile signal program modified to generate one component per ear. Subjects were trained on the interaural intensity difference task with pure tones presented at three center frequencies--500, 1000, and 2000 Hz, however because three of the four subjects were unable to perform the task, only 2000 Hz was used in this experiment. Each signal was

presented for 200 ms, with an inter-trial interval of approximately 500 ms.

C. Procedure

The experimental set-up for the interaural intensity difference (IID) experiment was similar to that used for the profile analysis experiment. Instrumentation for the interaural intensity difference stimuli is the same as that used for dichotic profile analysis.

For the IID task, a single frequency component was always presented to each ear in intervals one and four of the 2Q-2AFC task. An intensity increment, added to the component directed to the right ear (Channel 1), was presented in either interval two or three. This procedure is similar to the procedure used for the IID experiment by Nuetzel (1982, 1991).

Data collection was initiated for the interaural intensity difference task after the listeners had many months experience with the profile task for stimuli of both 500 ms in duration (pilot study) and 200 ms in duration.

Three conditions were tested for the IID task: a fixed condition with a 2000 Hz signal presented bilaterally, a roving level condition with a 2000 Hz signal presented bilaterally; and a roving level condition a 2000 Hz signal presented to the right ear and a 2350 Hz signal presented to the left ear. A 6 dB intensity increment was added to the right ear stimulus in all conditions. This intensity increment was selected since in the pilot dichotic profile analysis

studies, subjects capable of performing the profile analysis task with dichotic stimulus presentation were easily able to discriminate a 6 dB increment (Whitelaw, Hsu, and Feth, 1991).

III. RESULTS

Interaural intensity difference results presented here were obtained after subjects had practiced an average of 3500 trials. Gradual improvement in listener performance was not observed with practice for three of the four subjects (S1, S2, and S3).

Interaural intensity discrimination results are presented for individual subjects in Table 2. Inspection of these results demonstrates that two of the subjects (S1 and S4) were able to achieve threshold for the fixed listening condition, while the other two subjects (S2 and S3) did not. Of the two subjects able to perform the task in the fixed level condition, only one of them (S4) continued to demonstrate good performance in the roving level conditions while the other subject's performance decreased to chance. It is of interest to note that the subject who demonstrated the superior performance in the interaural intensity difference task (S4) is the subject who was unable to achieve threshold performance in the dichotic profile analysis task.

Overall, three of the four subjects who participated in this study were more sensitive to diotic profile presentations than for equivalent dichotic presentations. The fourth subject was unable to perform the dichotic profile task, but

was able to perform the task in the diotic condition. Statistical analysis confirmed that the diotic-dichotic differences were significant, however center frequency was not. Only one of the four subjects was capable of performing the IID task at a level substantially above chance, and that subject was the only subject who could not perform the dichotic profile task.

IV. GENERAL DISCUSSION

The results obtained in this study demonstrate diotic versus dichotic differences similar to those reported by Fantini et al. (1989) for a monotic-dichotic comparison. In their study, which used only a 2000 Hz center frequency, a 3 dB difference was reported between monotic and dichotic results, with the thresholds found to be 12 dB and 9 dB lower than the reference condition for the monotic and dichotic conditions, respectively. The 2.3 dB difference observed in the present study between diotic and dichotic thresholds at 2000 Hz is consistent with Fantini et al., despite methodological differences between these studies.

These results would appear to be substantially different from results reported by Green and Kidd (1983) and Bernstein and Green (1987). However, the differences observed may be related to disparity in reporting of the results. Many studies on profile analysis have expressed threshold results as the signal level relative to the single component level in the stimulus profile, including those by Green and Kidd

(1983) and Bernstein and Green (1987). In order to compare their results directly with the results of these two earlier studies, Fantini et al. (1989) converted the signal thresholds to $I + \Delta I$ values, then calculated masking release thresholds. Even when these converted results are considered, the dichotic release from masking continues to be less than that observed in the monotic condition in both studies. For Green and Kidd (1983), the difference between monotic and dichotic thresholds is 6 dB, which Fantini et al. contend may be comparable to those obtained in their study.

When Bernstein and Green's (1987) results were converted to masking release thresholds, a minimal masking release (2.7 dB) was noted for the dichotic listening condition, with a 7 dB difference noted between monotic and dichotic thresholds. The overall stimulus level used in their study was higher than that used by Green and Kidd (1983), Fantini et al. (1989), or the present study, which may have influenced the thresholds obtained.

When discussing profile analysis thresholds obtained for individual listeners, the "level detection limit" for the profile task must be considered. Green (1988) noted that the level limit might be the size of the intensity increment at which the listener is making a discrimination decision based on absolute intensity differences rather than spectral shape. The level limit depends on the rove range over which the stimulus is varied. Green (1988) has offered the value of

.2346 times the width of the roving range as the level detection limit for a 2 AFC task. In the present study, the level limit is approximately 5 dB. Several of the dichotic thresholds obtained in the present experiment (6 dB at 500 Hz for S3 and 7 dB at 1000 Hz for S2) exceed the level detection limit. This might suggest that under these conditions, subjects were basing their detection decision on a change in intensity level within one critical band and not on differences in spectral shape.

There is evidence, however, to suggest that subjects were not merely using absolute level differences in order to detect the addition of the target increment. Inspection of the psychometric functions obtained from subjects in this study show them to be smooth, which might be interpreted to indicate that listeners are using a similar listening strategy for all intensity increment steps. If a discontinuity were noted in the psychometric function, with a considerable improvement noted in the region of the threshold, a change in listening strategy might be suspected.

The differences between diotic and dichotic performance on the profile analysis task in this study may suggest a different mechanism or cue available to the listener in the dichotic task. One model proposed to explain the profile analysis phenomena is the channel theory (Durlach, Braida, and Ito, 1986; Green, 1988). The channel theory postulates that the signal is analyzed by the ear with a bank of

independent filters. The output of each filter is transformed so that the pressure level obtained from each channel is estimated, based on a nonlinear transformation process and temporal integration. The estimate is biased by random, internal noise. This model asserts that the listener must determine if the output of a particular channel "stands out" because of the addition of a more intense signal or because of this internal noise. Since the filters are assumed to be independent of each other, performance should be unaffected by moving some of the signal components to the other ear. That is, the same performance would be anticipated for both diotic and dichotic listening conditions. This does not appear to be the case in the present study.

Comparing the slopes of the psychometric functions of dichotic to diotic results may also provide additional support for a different mechanism for dichotic profile analysis. If the same mechanism were responsible for both diotic and dichotic profile analysis, the slope of the function for the dichotic results should be similar to that of the diotic results, merely shifted to the right. This hypothesis is based on previous research which suggested that diotic thresholds were lower than thresholds obtained from dichotic stimulus presentation. From inspection of the psychometric functions, the dichotic results are not only shifted to the right but the slope is also flatter than for the diotic condition. This observation is confirmed by the statistical tests of

difference for the slopes of the lines for all subjects at 500 and 1000 Hz and for one subject at 2000 Hz. This lends support to the notion that the dichotic cue may be different from the cues used to discriminate the diotic profile.

The lack of a significant frequency effect might support the concept that subjects are using something other than an intensity cue for extracting the profile. A wealth of research has indicated that a dual mechanism exists for processing binaural information. For stimuli below 1500 Hz, binaural processing appears to be based primarily on interaural time cues. For stimuli above 1500 Hz, both interaural time and interaural intensity cues appear to be important in the interpretation of the signal (Jeffress, 1971). Performance on the IID task was found to be consistent with this notion. IID detection was found to improve for tones presented at 2000 Hz and above (Nuetzel, 1982). It was anticipated that if dichotic profile analysis were based merely on an interaural intensity difference, lower discrimination thresholds would be observed for higher center frequencies. This result was not observed for the present study.

The results of the subject (S4) who could not achieve threshold in the dichotic condition require further comment. In many early studies in profile analysis, listeners who failed to achieve threshold after a number of trials were excluded from the experiment (Henn and Turner, 1990). This

subject was included in the present study to support the notion that profile analysis may not be a "universal effect" (Henn and Turner, 1990). This subject may be using a different strategy for listening for spectral shape information, which might be representative of a strategy used by a subgroup of listeners for these types of tasks.

In order to extract the profile from the signal used in the present experiment, it appears that listeners require information be presented to both ears simultaneously. However, alternative explanations to combination of dichotic information in order to achieve an integrated spectral stimulus must be considered. Instead of combining the information from the two ears into a fused signal, listeners might be able to simultaneously monitor the intensity level in each ear and compare the absolute intensity levels between the two ears, rather than making the decision based on spectral shape (Mason, 1991).

Independent intensity level comparisons are the basis for the interaural intensity difference (IID) experiment. It has been postulated that listeners can use an intensity comparison between the two ears to detect very small intensity increments, particularly with tonal stimuli in the 2000 to 3000 Hz region. Subjects in this study were tested first for the conditions which Nuetzel (1982) reported maximum sensitivity for to interaural intensity differences.

It is somewhat surprising that the three subjects who were able to perform the dichotic profile task were unable to achieve greater than chance performance for a roving level IID task after more than 3000 trials. Based on these results, it would appear that listeners may not be using an interaural absolute level detection strategy for the profile task. If this were the case, the same subjects should have easily demonstrated comparable performance on the IID task to that obtained for the dichotic profile analysis task.

Only one subject (S4) demonstrated stable and consistently good performance on the IID task. She is also the only subject who did not achieve threshold on the dichotic profile analysis task. For the IID task, her performance was not substantially degraded by the addition of the roving level condition after training on the fixed level condition; in the roving level condition, the performance for the other three subjects fell to chance. S4 reported that she did not hear a 'fused' image for the profile task and that she was not using a 'movement' cue for the IID task. The results obtained on this subject, along with her reported subjective impressions, suggest that she may be monitoring intensity at each ear individually, then comparing the waveforms at both ears. Since she was unable to achieve threshold for dichotic profile task, it would appear that this strategy is ineffective for discriminating spectral shape.

The findings of the present study suggest that most subjects are able to extract a profile from the dichotic signal presentation, although the effect is not as robust as when the signal is presented dichotically. This ability to detect spectral shape information from a signal presented to both ears may be interpreted as evidence to support a central mechanism for profile analysis effects.

A parallel between the dichotic profile and other auditory tasks thought to be centrally mediated may exist. For example, a 'residue' pitch can be heard when there is no possibility of interaction of the auditory signal at the level of the cochlea (Houtsma and Goldstein, 1972; Moore, 1988). Listeners were asked to recognize melodies corresponding to missing fundamentals. Listeners were able to extract the residue pitch in order to identify melodies when the signal components were presented dichotically, however monotic performance was superior to dichotic performance at high intensities. Therefore, despite the fact that the dichotic performance is inferior to monotic performance, the effect persists.

The superior performance for the monotic condition over the dichotic condition at high intensities has been attributed to the presence of combination tones which provide additional harmonics which serve to strengthen the pitch in this condition (Houtsma and Goldstein, 1972). Combination tones do not occur for sinusoids presented dichotically, and they

are much weaker when presented monotically at intensity levels near threshold. Therefore, the dichotic performance obtained for the residue pitch experiments may reflect the "true" performance of the auditory system, and the superior monotic performance may be an artifact of additional cues which are available due to non-linearities of the cochlea.

A similar argument could be applied to the profile analysis results obtained in the present study. It is possible that dichotic results better represent the listeners actual performance on the task. Monotic and diotic results might be enhanced by a frequency modulation (FM) cue, as suggested by Feth and Stover (1987). They suggested that interactions among the components of the profile analysis stimuli result in FM artifacts which are detected by the listener. These modulation cues, which are available to the listener in monotic and diotic profile analysis tasks, artificially enhance listener performance. These cues are not salient in the dichotic listening situation, thus the "poorer" performance.

V. SUMMARY

The major findings of this study are as follows:

- 1) The results of these experiments suggest that most subjects are able to perform the dichotic profile analysis task. This observation, in addition to significant differences in slopes of the psychometric functions between the diotic and dichotic results, might suggest that a different mechanism is

available for processing of spectral shape information presented dichotically.

2) A significant frequency effect was not noted for either diotic or dichotic stimuli. This is particularly noteworthy in the dichotic condition, given that simple binaural intensity summation tasks demonstrate a significant frequency effect, with lower thresholds observed for higher frequency stimuli. Since a frequency effect was not observed for the results of the present study, a simple intensity summation for information presented to both ears is not suspected as the underlying explanation for the dichotic profile analysis results.

3) Although similarities may exist between dichotic profile analysis and other types of binaural listening tasks, profile analysis and interaural intensity differences do not appear to result from the same mechanism. Subjects who were capable of performing the dichotic profile task were unable to perform the IID task, even after considerable training. Conversely, the only listener in the present study who easily learned the IID task under the maximum performance conditions reported by Nuetzel (1982) was unable to perform the dichotic profile analysis task.

4) Significant individual differences in listener performance were noted for the profile analysis task. This is consistent with the results obtained in previous studies on complex auditory processing in general and specifically in other

profile analysis studies.

The results of the present study suggest that profile analysis is not mediated on only the peripheral level and that a central mechanism for extracting information from the signal may exist. These findings demonstrate lower dichotic thresholds than obtained in the majority of previous research in dichotic profile analysis, which may be related, in part, to methodological differences across studies.

REFERENCES

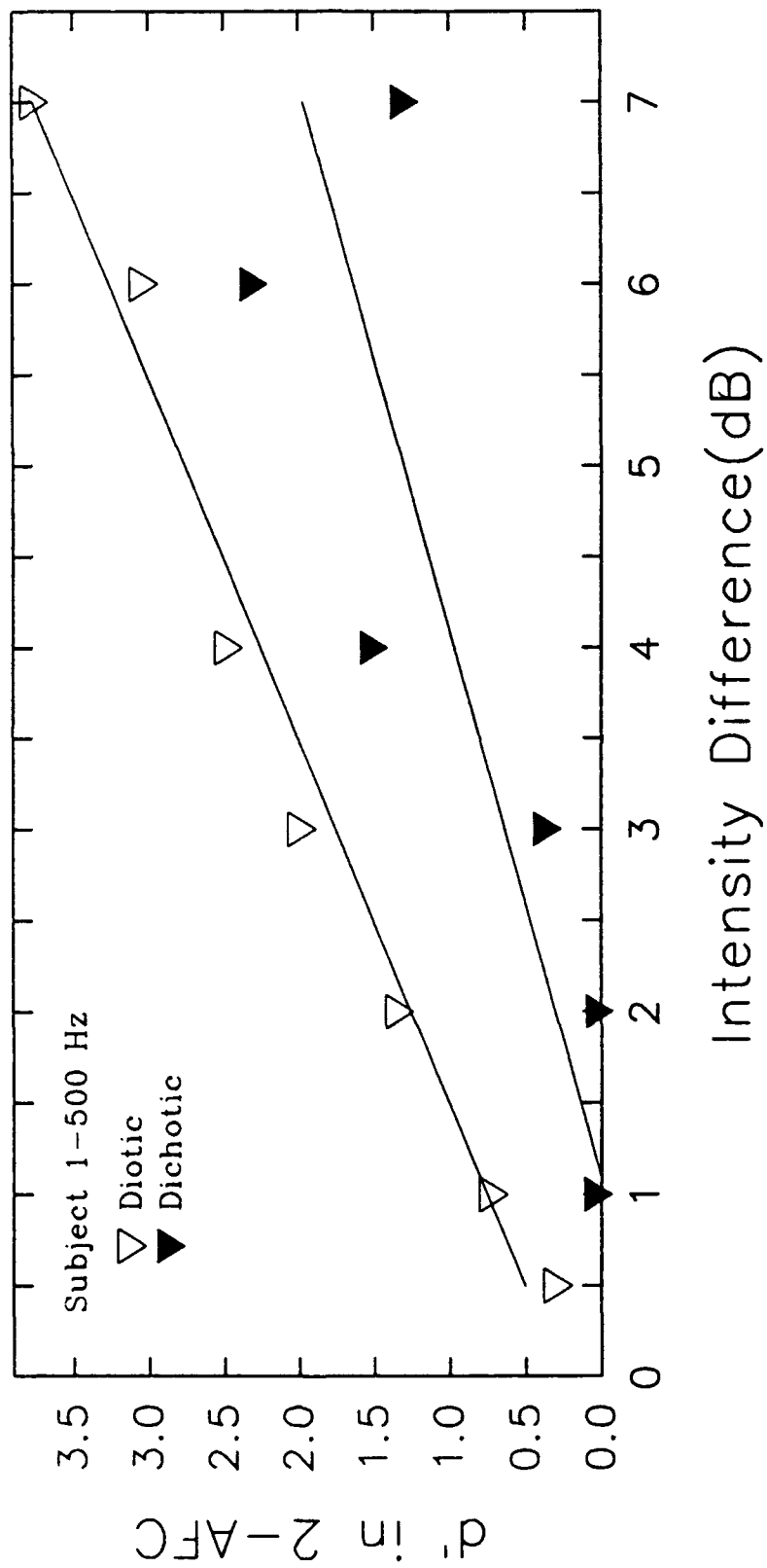
- Berliner, J.E. & Durlach, N.I. (1973). Intensity perception. IV. Resolution in roving-level discrimination. J. Acoust. Soc. Am., 53 (4), 1270-1287.
- Bernstein, L.R. (1990). Personal communication.
- Bernstein, L.R. & Green, D.M. (1987). The profile-analysis bandwidth. J. Acoust. Soc. Am., 81(6), 1888-1895.
- Cohen, J. & Cohen, P. (1983). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (2nd. Ed). Hillsdale, NJ: Lawrence Erlbaum.
- Durlach, N.I., Braida, L.D., & Ito, Y. (1986). Toward a model for discrimination of broadband signals. J. Acoust. Soc. Am., 80(1), 63-72.
- Dye, R.H. (1990). The combination of interaural information across frequencies: Lateralization on the basis of interaural delay. J. Acoust. Soc. Am., 88(6), 2159-2170.
- Fantini, D., Schooneveldt, G.P., & Moore, B.C.J. (1989). A comparison of comodulation masking release and profile analysis. J. Acoust. Soc. Am., 86 (Supplement 1), S10.
- Feth, L.L. & Stover, L.J. (1987). Demodulation process in auditory perception. In W.A. Yost & C.S. Watson (Eds.) Auditory Processing of Complex Sounds (p.76-86). Hillsdale, NJ: Lawrence Erlbaum
- Green, D. (1988). Profile Analysis: Auditory Intensity Discrimination. New York: Oxford University Press.
- Green, D. (1988). Audition: Psychophysics and Perception. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey, and R.D. Luce (Eds.) Stevens' Handbook of Experimental Psychology (2nd Ed.) Volume 1--Perception and Motivation (p. 327-376). New York: John Wiley and Sons.
- Green, D.M. & Kidd, G., Jr. (1983). Further studies of auditory profile analysis. J. Acoust. Soc. Am., 73(4), 1260-1265.

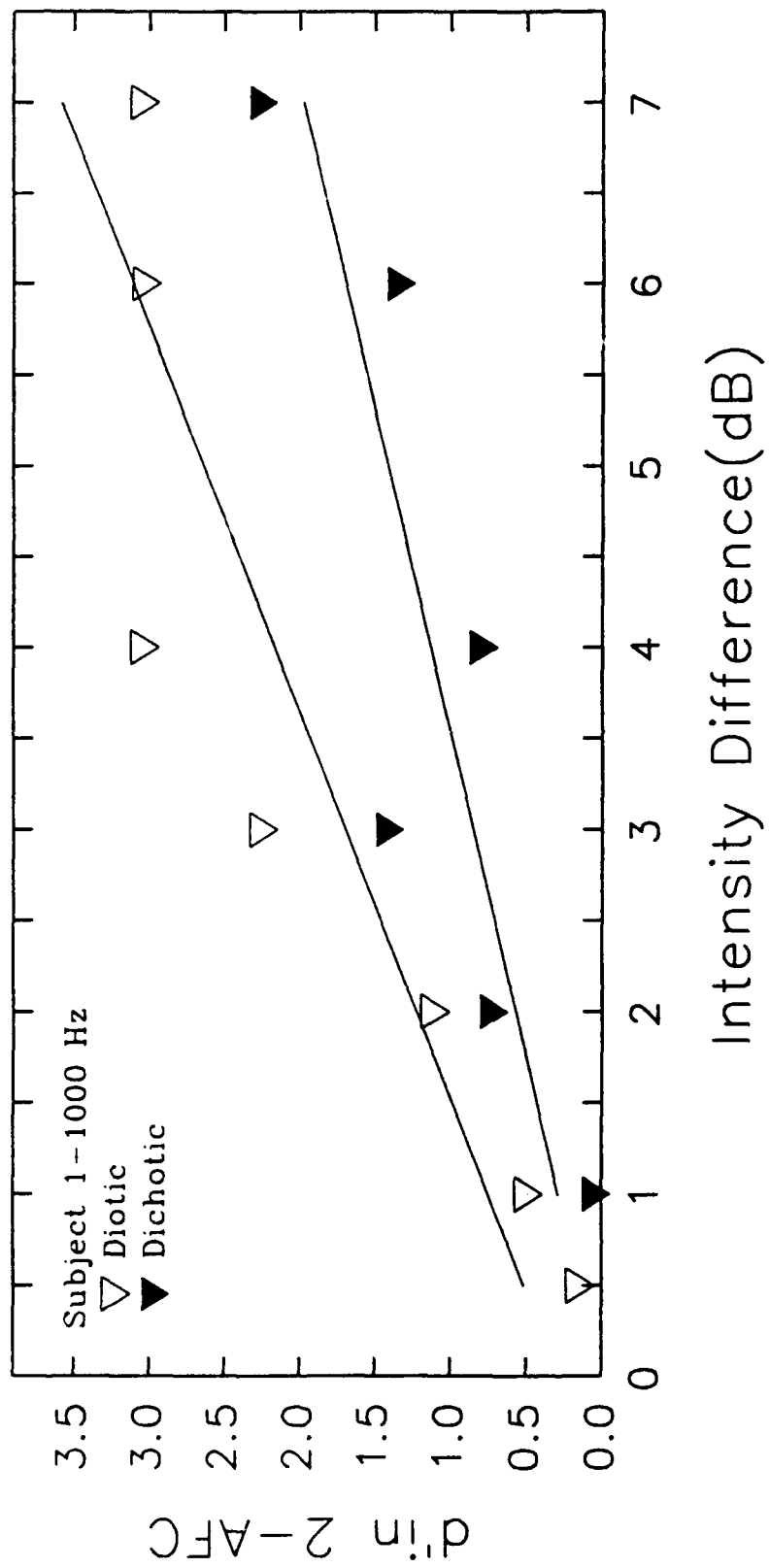
- Henn, C.C. & Turner, C.W. (1990). Pure-tone increment detection in harmonic and inharmonic backgrounds. J. Acoust. Soc. Am., 88(1), 126-131.
- Houtsma, A.J.M. & Goldstein, J.L. (1972). The central origin of the pitch of complex tones: Evidence from musical recognition. J. Acoust. Soc. Am., 51(2), 520-529.
- Hsu, C.Y. & Feth L.L. (1992). Using the Ariel DSP-16 as a signal generator for psychoacoustics experiments. Paper submitted to Behavioral Research Methods, Computers, and Instrumentation.
- Jeffress, L.A. (1971). Detection and lateralization of binaural signals. Audiology, 10, 77-84.
- Julesz, B. (1971). Foundations of Cyclopean Perception. Chicago: The University of Chicago Press.
- Krajewski, L.T. & Ritzman, L.P. (1990). Operations Management: Strategy and Analysis(2nd ed.). Reading, MA: Addison-Wesley Publishing.
- Mason, C.R. (1991). Personal communication.
- McNichol, D. (1972). A Primer of Signal Detection Theory. London: George Allen and Unwin.
- Moore, B.C.J. (1988). An Introduction to the Psychology of Hearing (2nd Ed.). San Diego, CA: Academic Press.
- Nuetzel, J.M. (1982). Sensitivity in interaural intensity differences to tones and noises as measured with a roving level paradigm. J. Acoust. Soc. Am., 71 (Supplement 1), S47.
- Nuetzel, J.M. (1991). Personal communication.
- Whitelaw, G.M., Hsu, C.Y. & Feth, L.L. (1991). Dichotic profile analysis. J. Acoust. Soc. Am., 89(4), 4PP3.

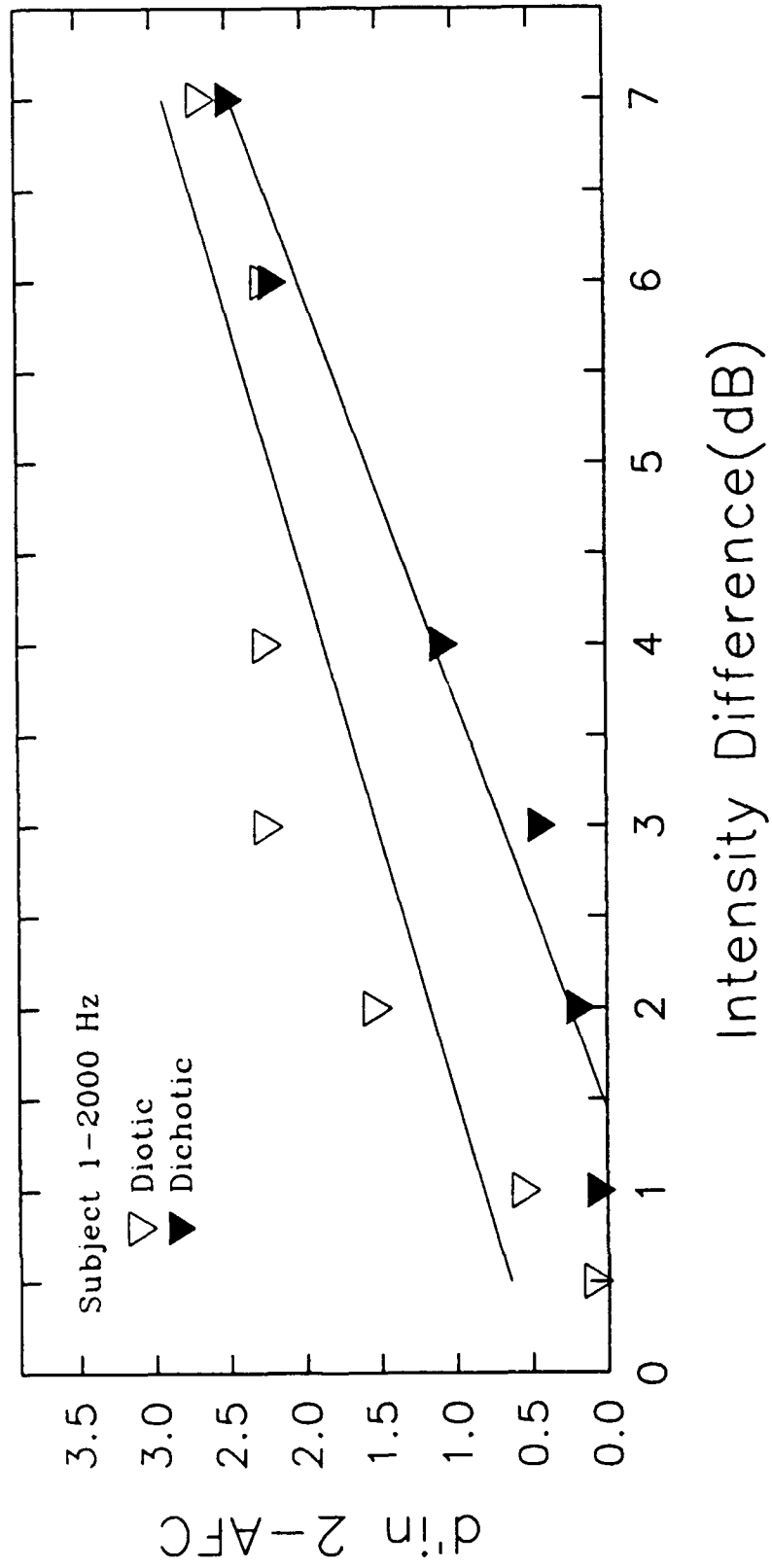
This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, Under Grant AFOSR089-0227.

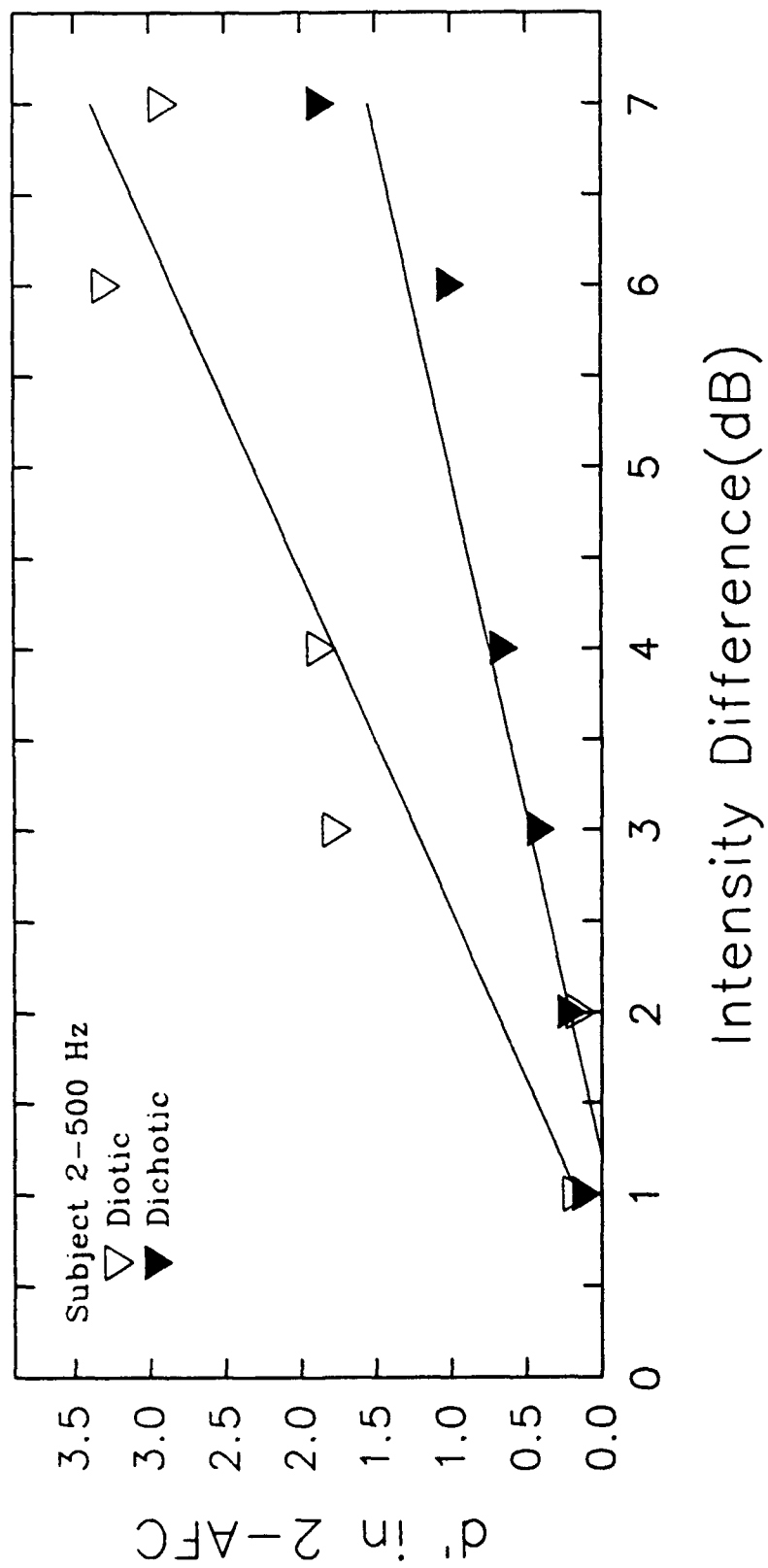
Figure 1: Dichotic stimulus configurations a) Green and Kidd, 1983; Bernstein and Green, 1987 b) Fantini, Schooneveldt, and Moore, 1989 c) Whitelaw, Hsu, Feth, 1991.

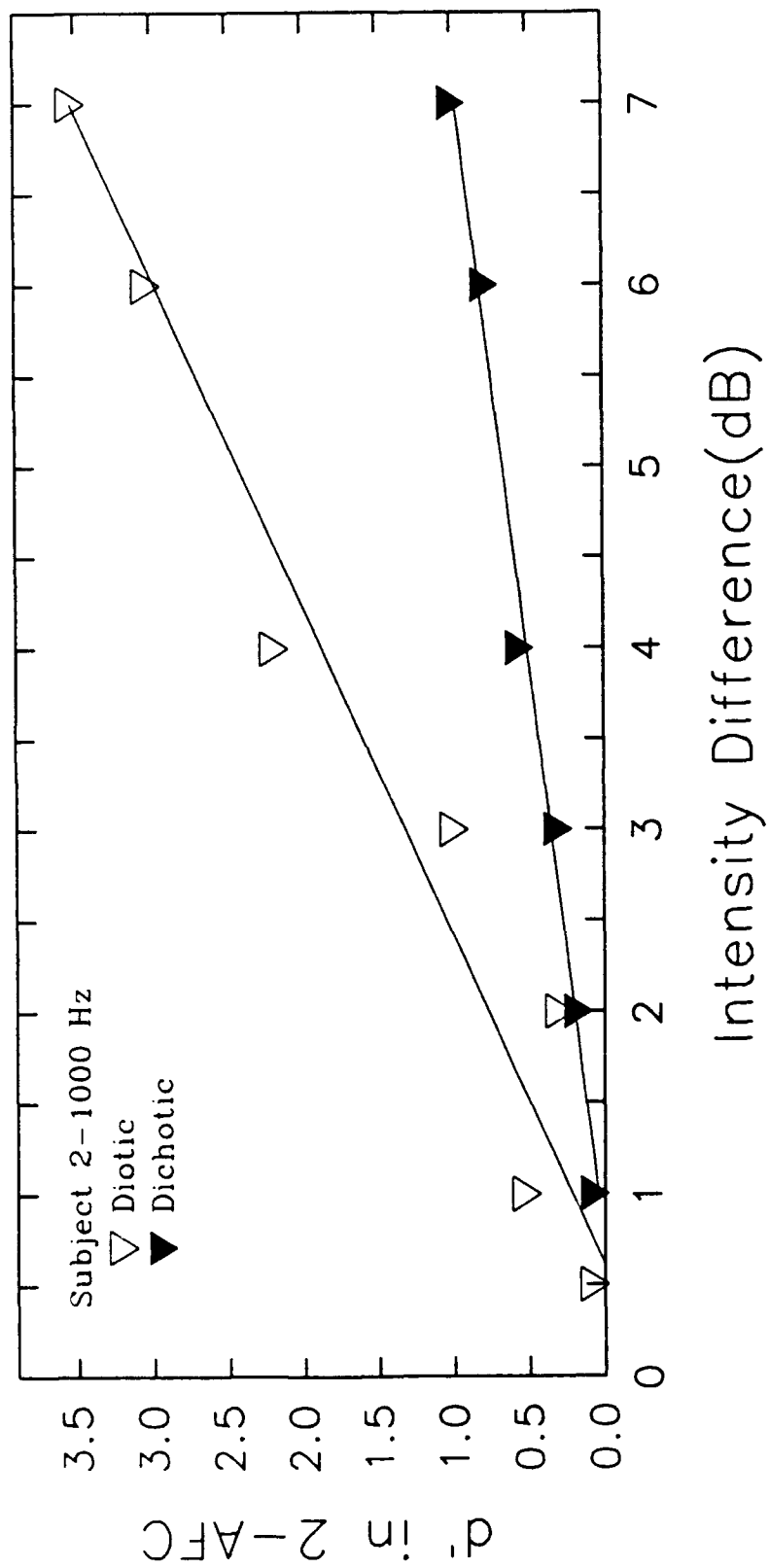
Figures 2a-d: Psychometric functions for individual subjects for
diotic and dichotic profile analysis performance.

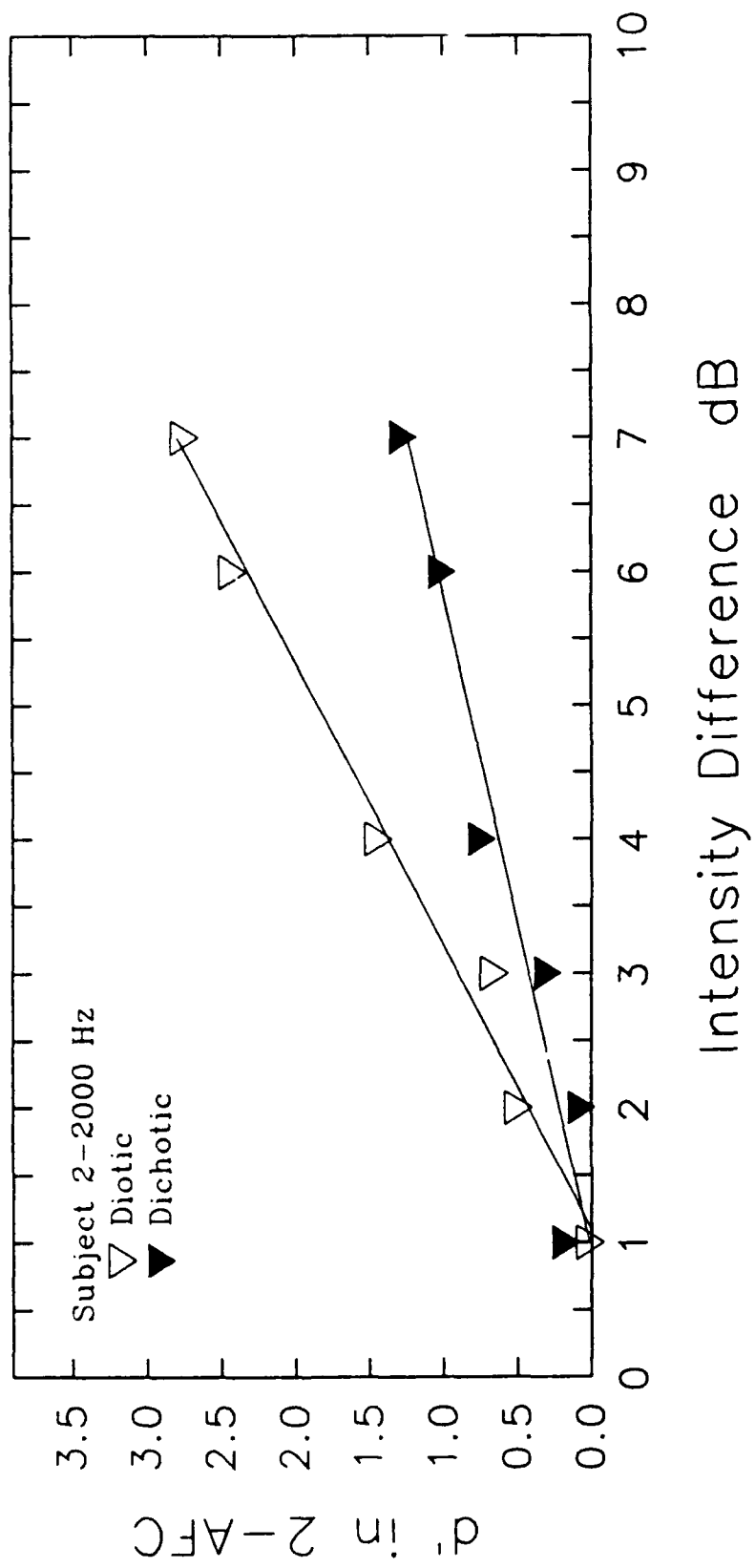


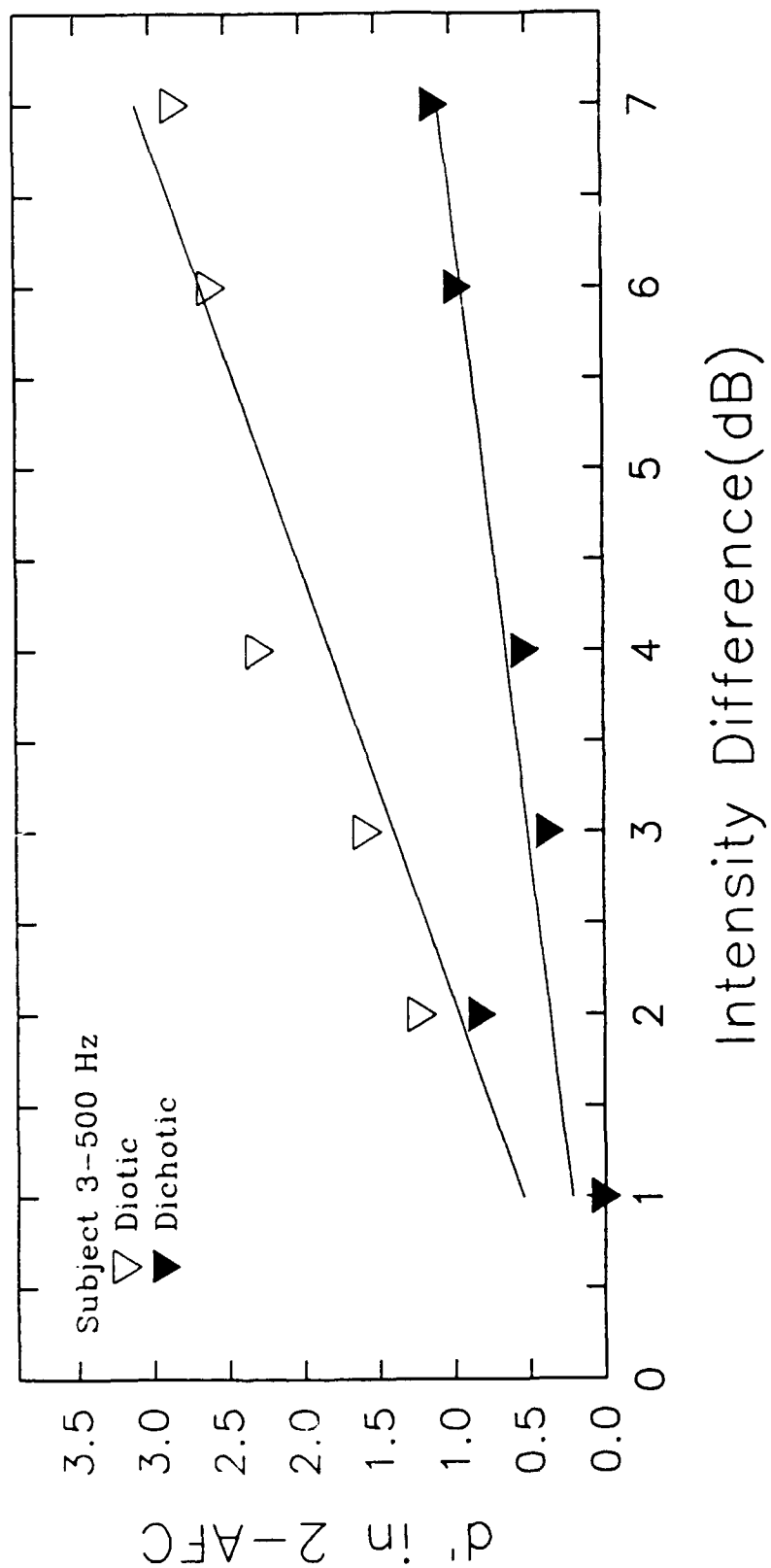


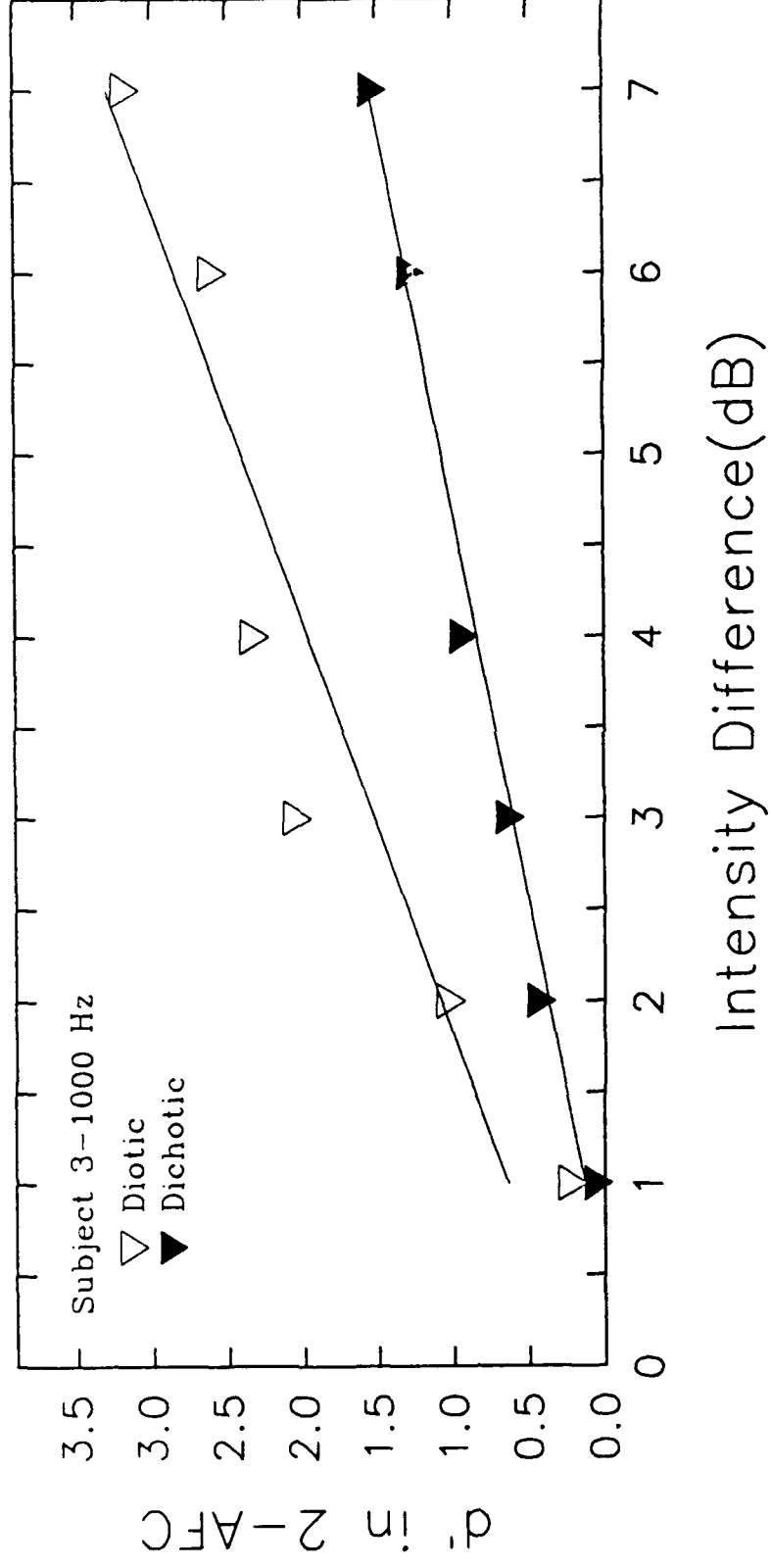


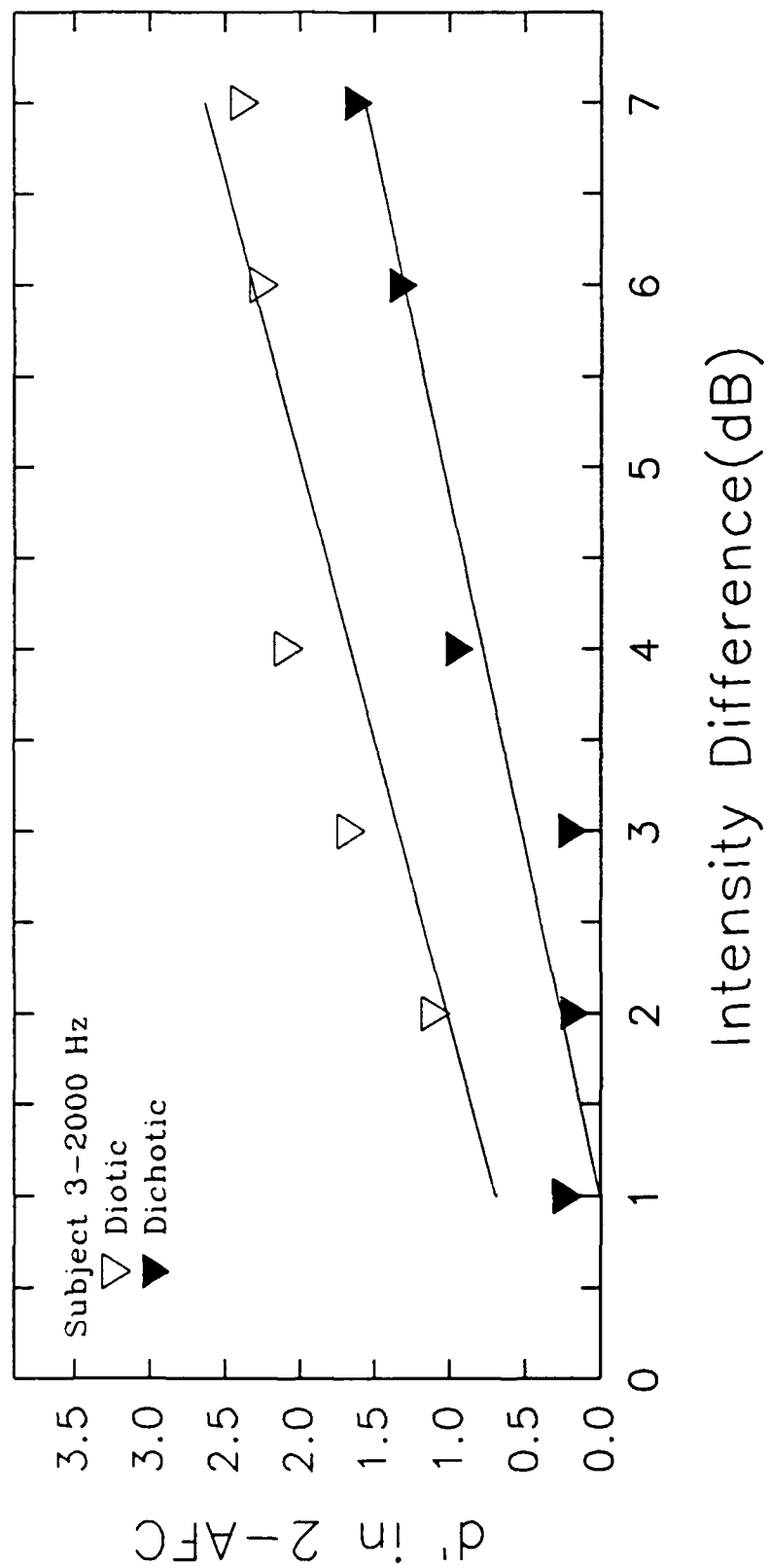


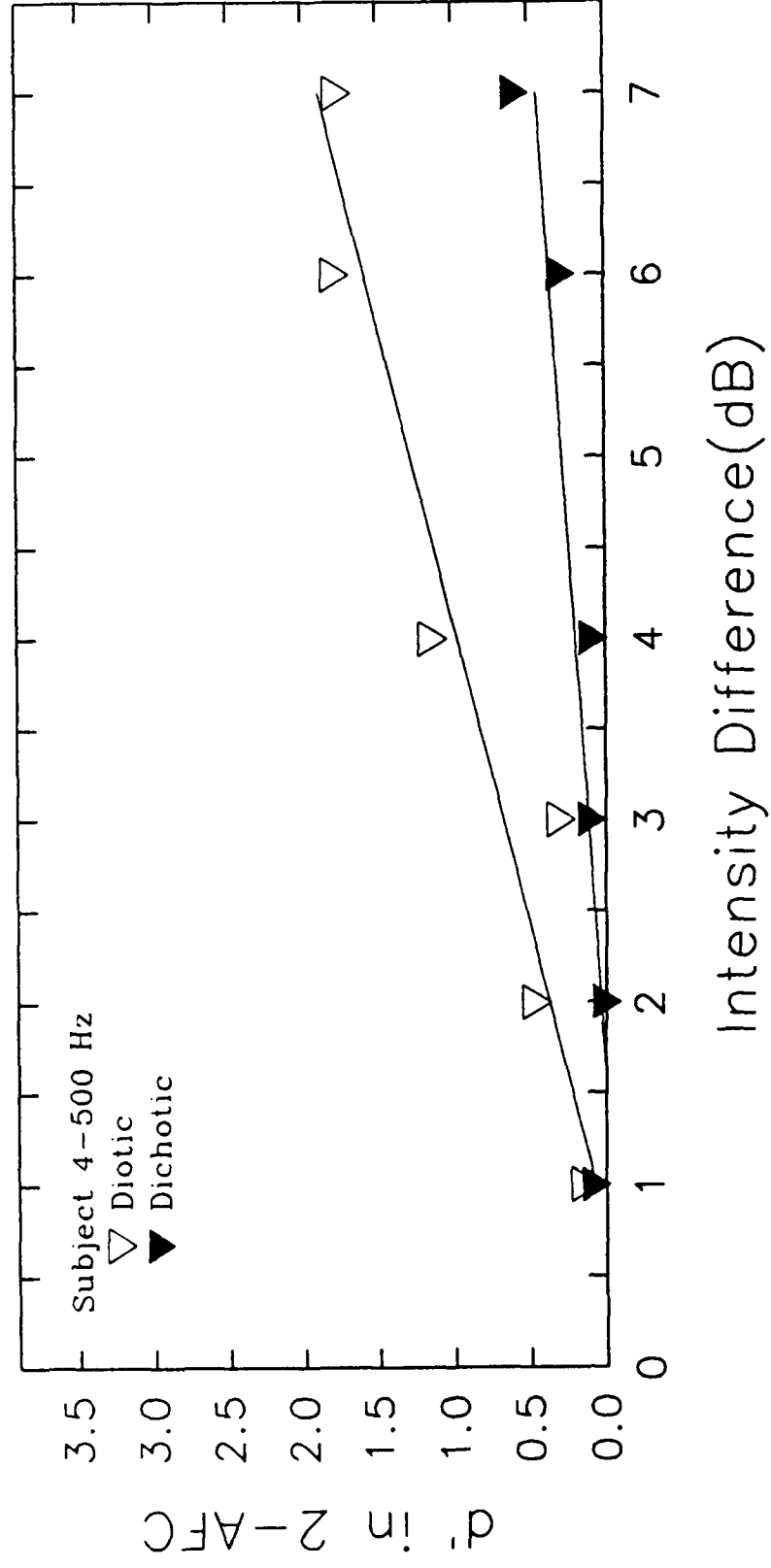


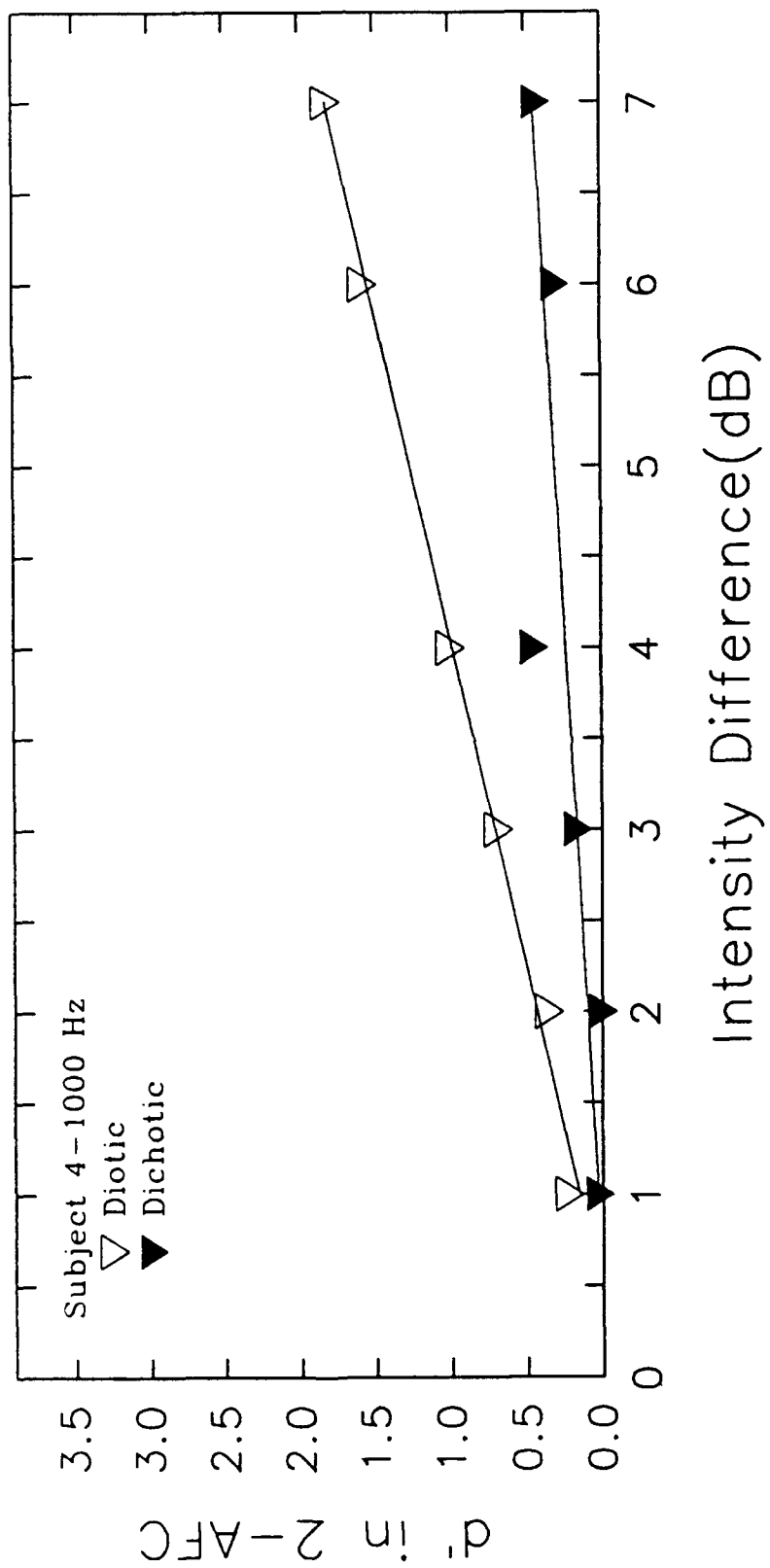


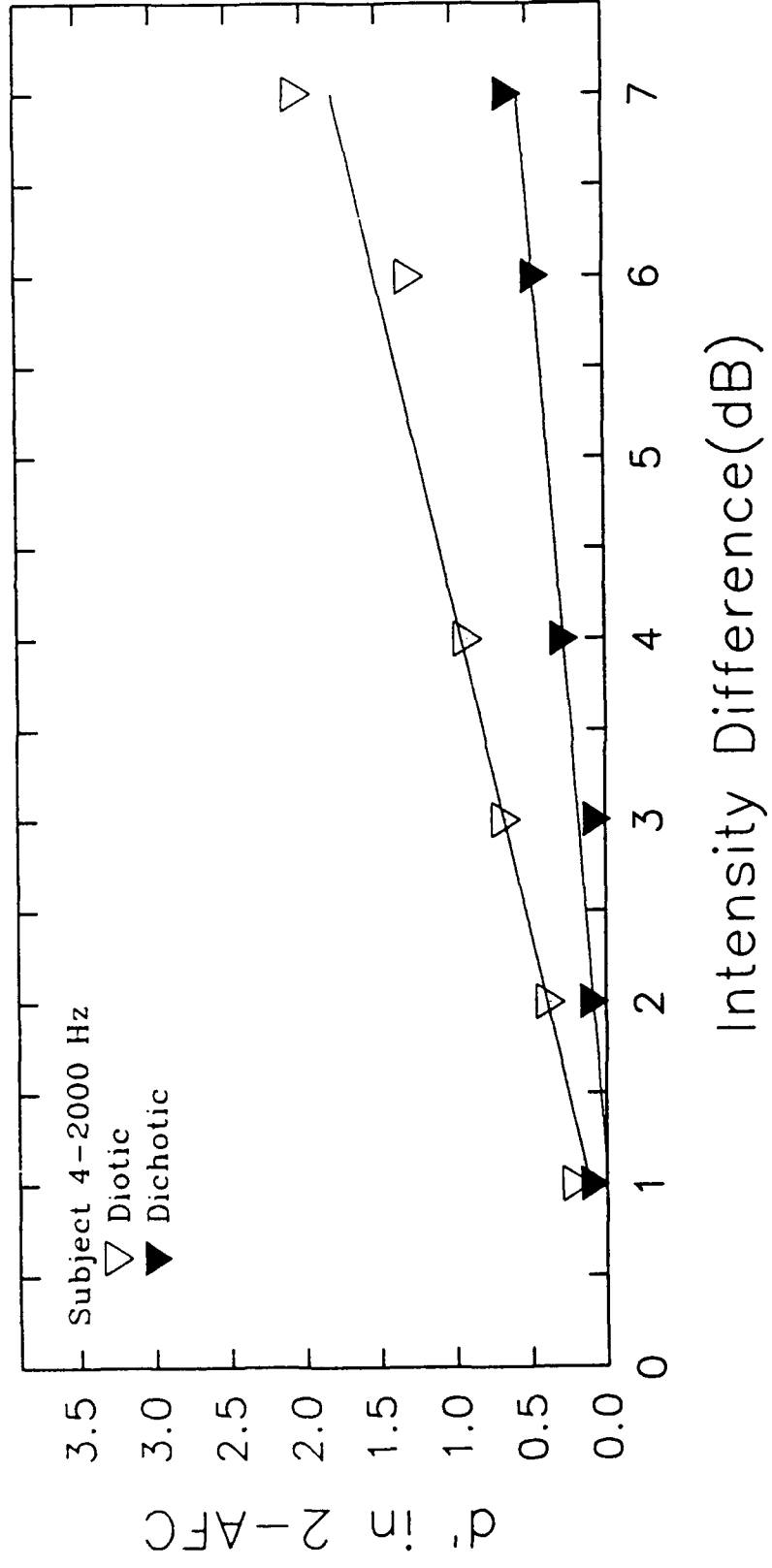




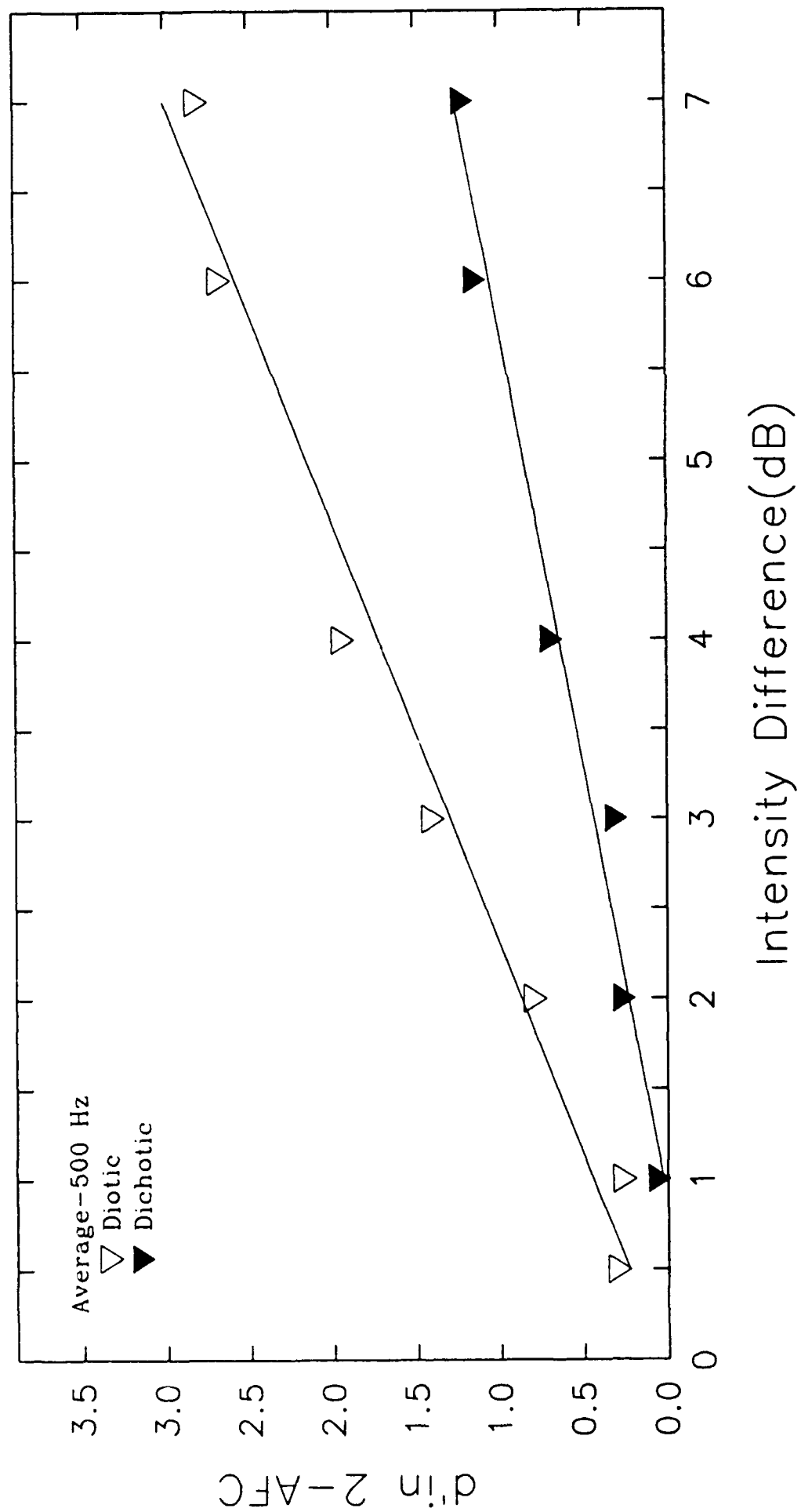


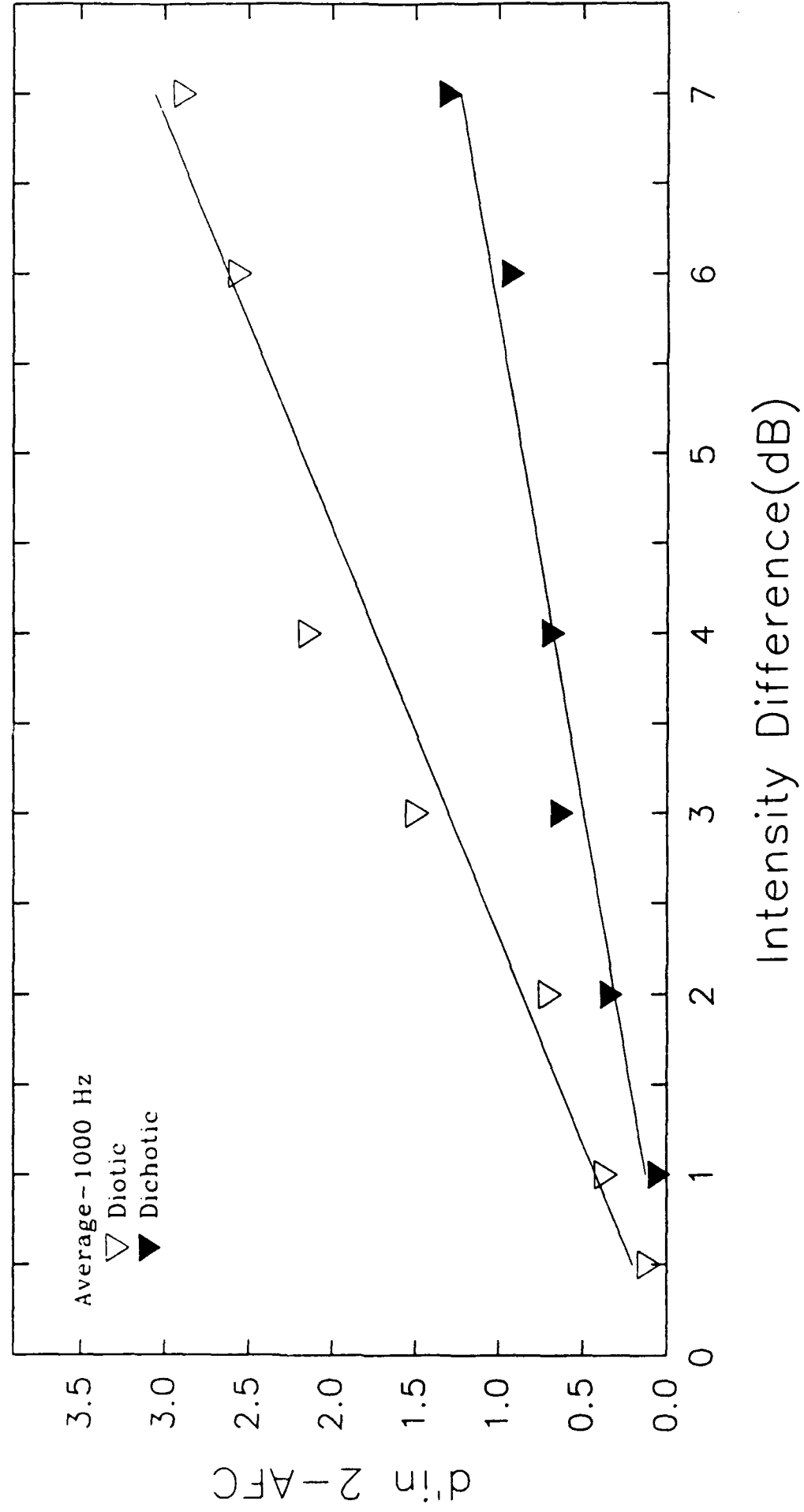






Figures 3a-c: Psychometric function for group mean performance for 500, 1000, and 2000 Hz.





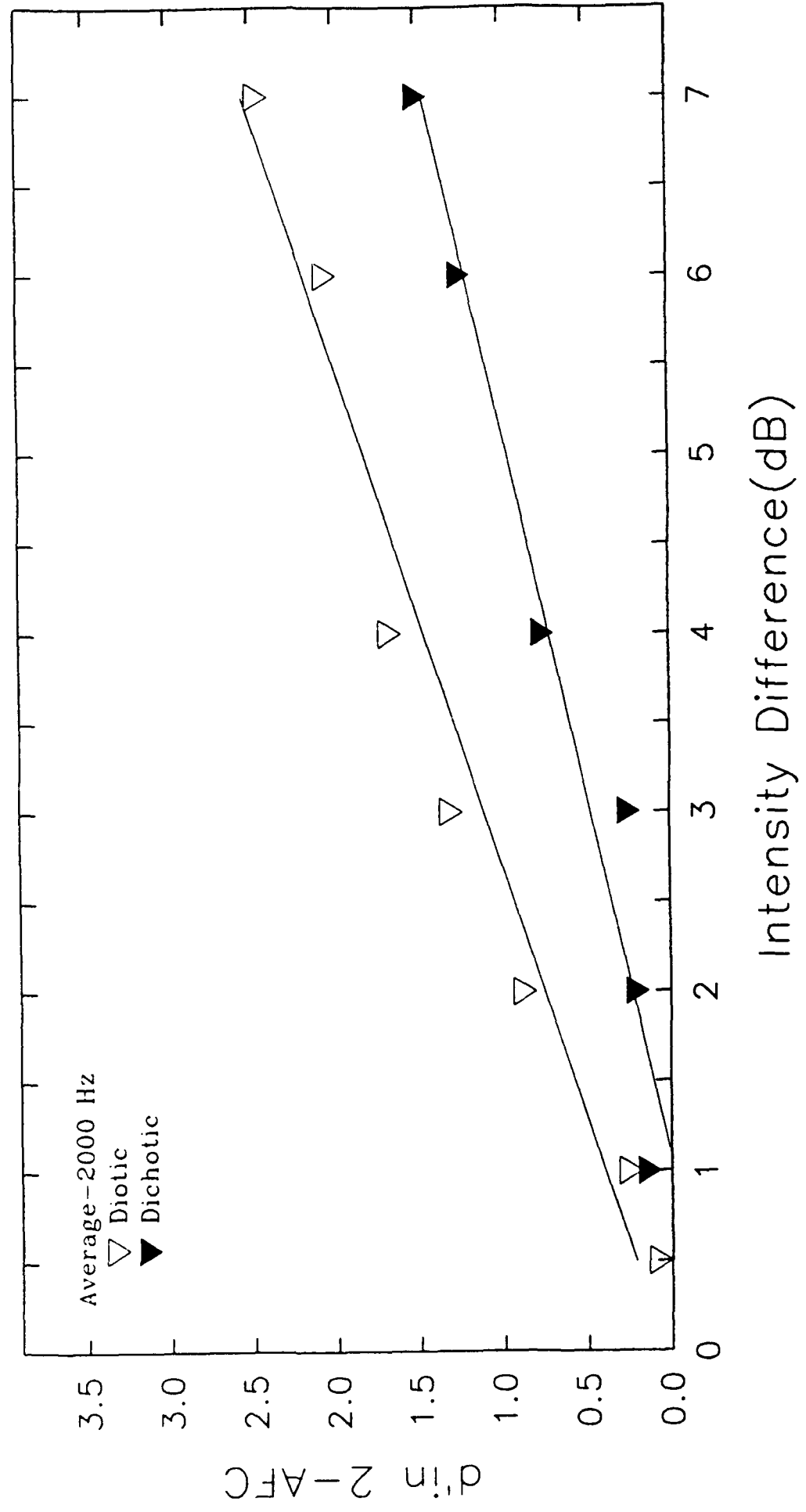


TABLE 1

DISCRIMINATION THRESHOLDS FOR 2Q-2AFC EXPRESSED IN dB

500 Hz

<u>Subject</u>	<u>Diotic</u>	<u>Dichotic</u>
1	1.5	4.0
2	2.5	4.5
3	2.0	6.0
4	4.0	N/A

1000 Hz

<u>Subject</u>	<u>Diotic</u>	<u>Dichotic</u>
1	1.5	3.0
2	2.5	7.0
3	2.0	4.5
4	4.0	N/A

2000 Hz

<u>Subject</u>	<u>Diotic</u>	<u>Dichotic</u>
1	1.5	3.5
2	3.0	5.5
3	2.0	4.5
4	4.0	N/A

TABLE 2

PERCENT CORRECT SCORES FOR INTERAURAL INTENSITY DIFFERENCES

Fixed condition (2000 Hz presented to both ears, 6 dB increment)

<u>Subject</u>	<u>Percent correct</u>
1	86.3
2	70.3
3	56.7
4	93.3

Roving condition (+/-10 dB) (2000 Hz presented to both ears, 6 dB increment)

<u>Subject</u>	<u>Percent correct</u>
1	47.3
2	57.7
3	58.0
4	90.7

Roving condition (+/-10 dB) (2000 Hz presented to right ear and 2350 Hz presented to left ear, 6 dB increment)

<u>Subject</u>	<u>Percent correct</u>
1	52.7
2	61.0
3	52.3
4	84.3

Envelope Weighted Average of Instantaneous Frequency

Signal, $s(t) = e(t) \cos[\theta(t)]$

$e(t)$ is the instantaneous envelope and
 $\theta(t)$ is the instantaneous phase.

Instantaneous frequency, $f(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt}$

EWAIF of signal $s(t)$, $0 \leq t \leq T$,

$$EW_s = \frac{\int_0^T e(t) f(t) dt}{\int_0^T e(t) dt}$$

Intensity Weighted Average of Instantaneous Frequency

IWAIF of signal $s(t)$, $0 \leq t \leq T$,

$$IW_s = \frac{\int_0^T e^2(t) f(t) dt}{\int_0^T e^2(t) dt}$$

EWAIF Envelope is the weighting function.

IWAIF Square of envelope is the weighting function.

EWAIF and IWAIF values are highly correlated.

IWAIF in the Frequency Domain

IWAIF of signal $s(t)$, $0 \leq t \leq T$,

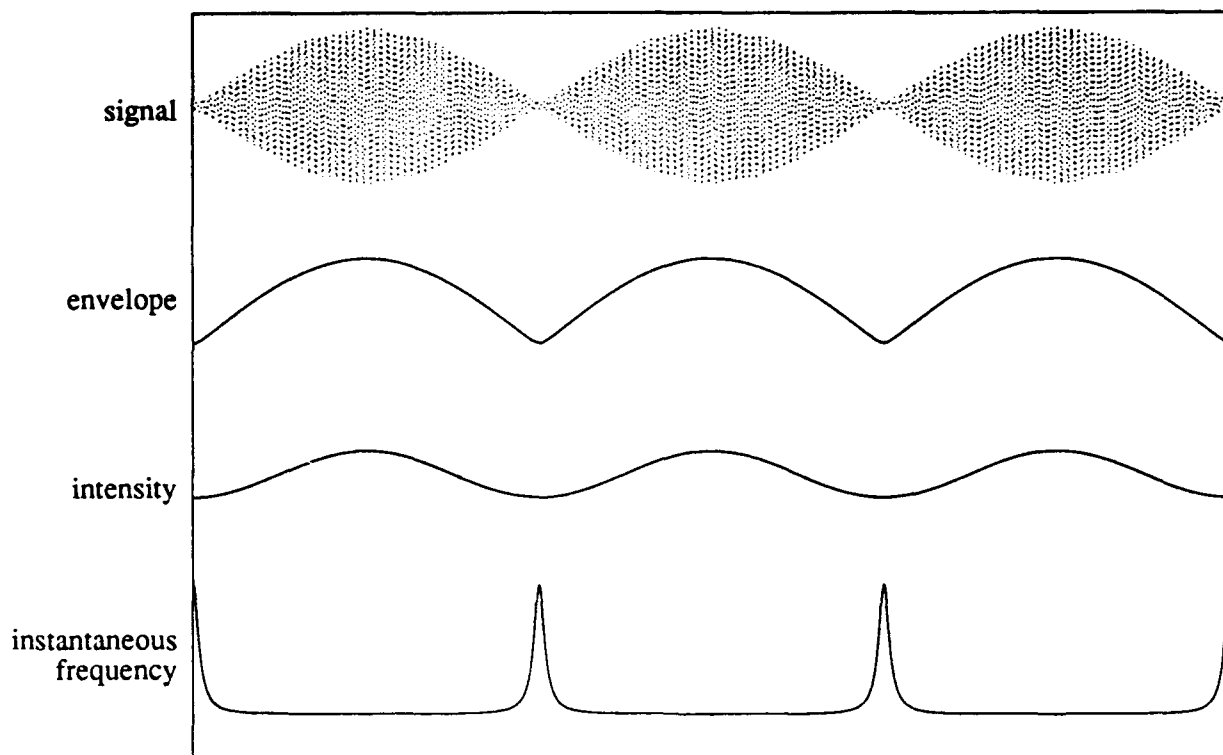
$$IW_s = \frac{\int_0^\infty \omega |S(\omega)|^2 d\omega}{\int_0^\infty |S(\omega)|^2 d\omega}$$

$S(\omega)$ is the Fourier transform of $s(t)$.

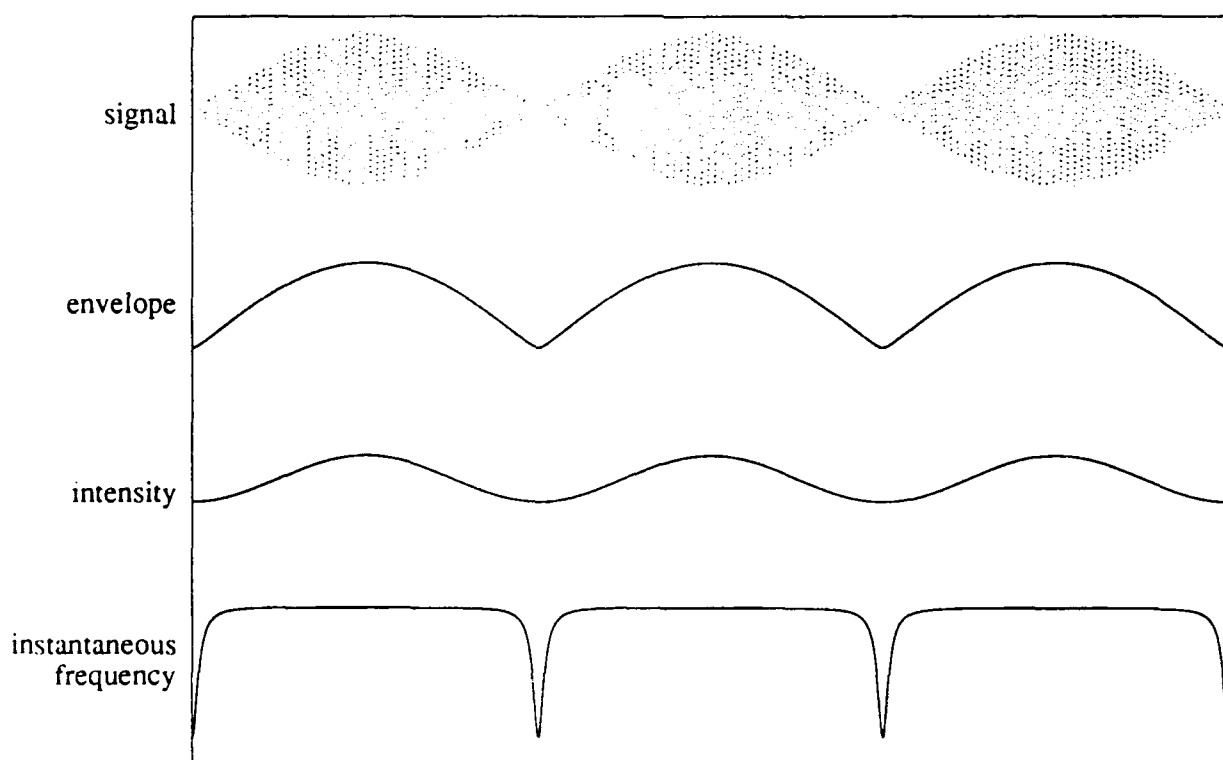
This representation leads to

- A fast and easy computation of the IWAIF using FFT algorithms.
- A more tractable model with the filterbank introduced.

Henceforth IWAIF of signals is calculated in the frequency domain.



(1000 Hz, 71 dB) (1020 Hz, 70 dB)



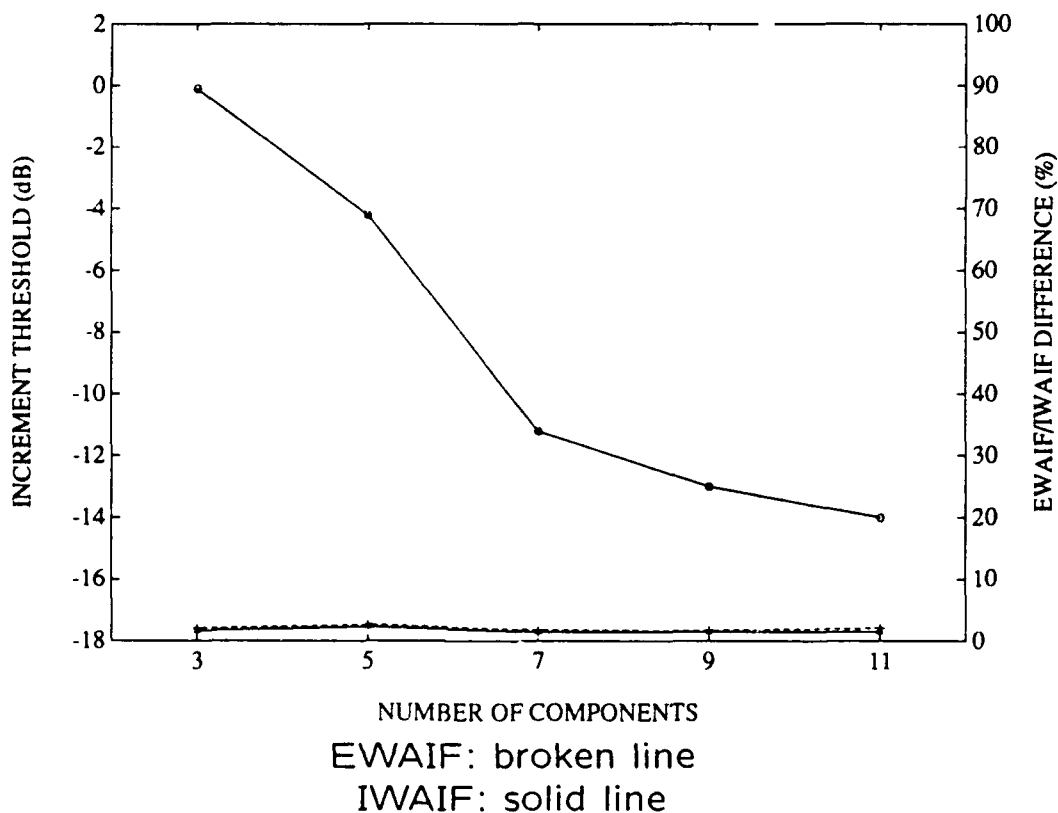
(1000 Hz, 70 dB) (1020 Hz, 71 dB)

Figure 1. Sampled signal, envelope and intensity functions, and instantaneous frequency fluctuations for a typical complimentary pair of Voelcker tones.

Comparision of IWAIF and EWIAIF

signal	EWIAF	IWAIF
(1000 Hz, 71 dB) (1020 Hz, 70 dB)	1007.59	1008.82
(1000 Hz, 70 dB) (1020 Hz, 71 dB)	1012.41	1011.18

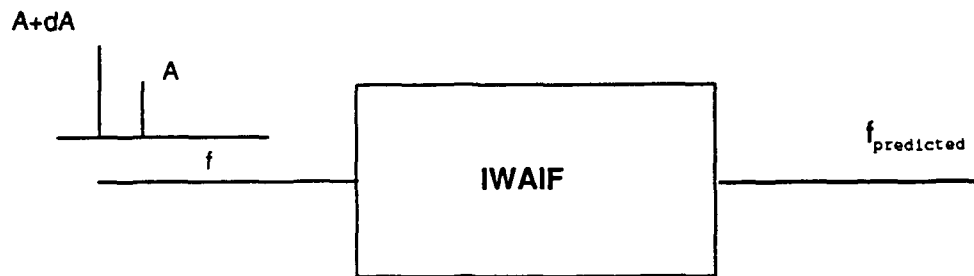
EWIAF/IWAIF values for complementary Voelcker signal pairs



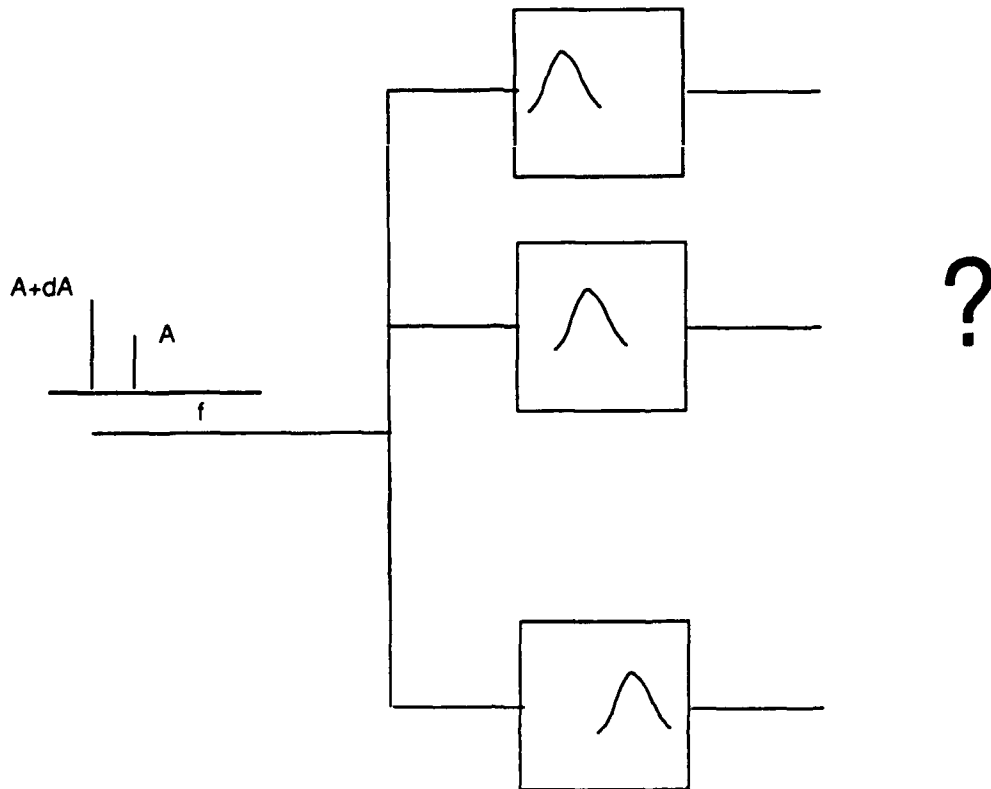
Profile signals (Green, Mason, & Kidd 1984)

Multichannel IWAIF

Problem: Given a complex signal, we know how to compute it's IWAIF. How do we get a measure similar to the IWAIF that accounts for the filtering by the basilar membrane ?

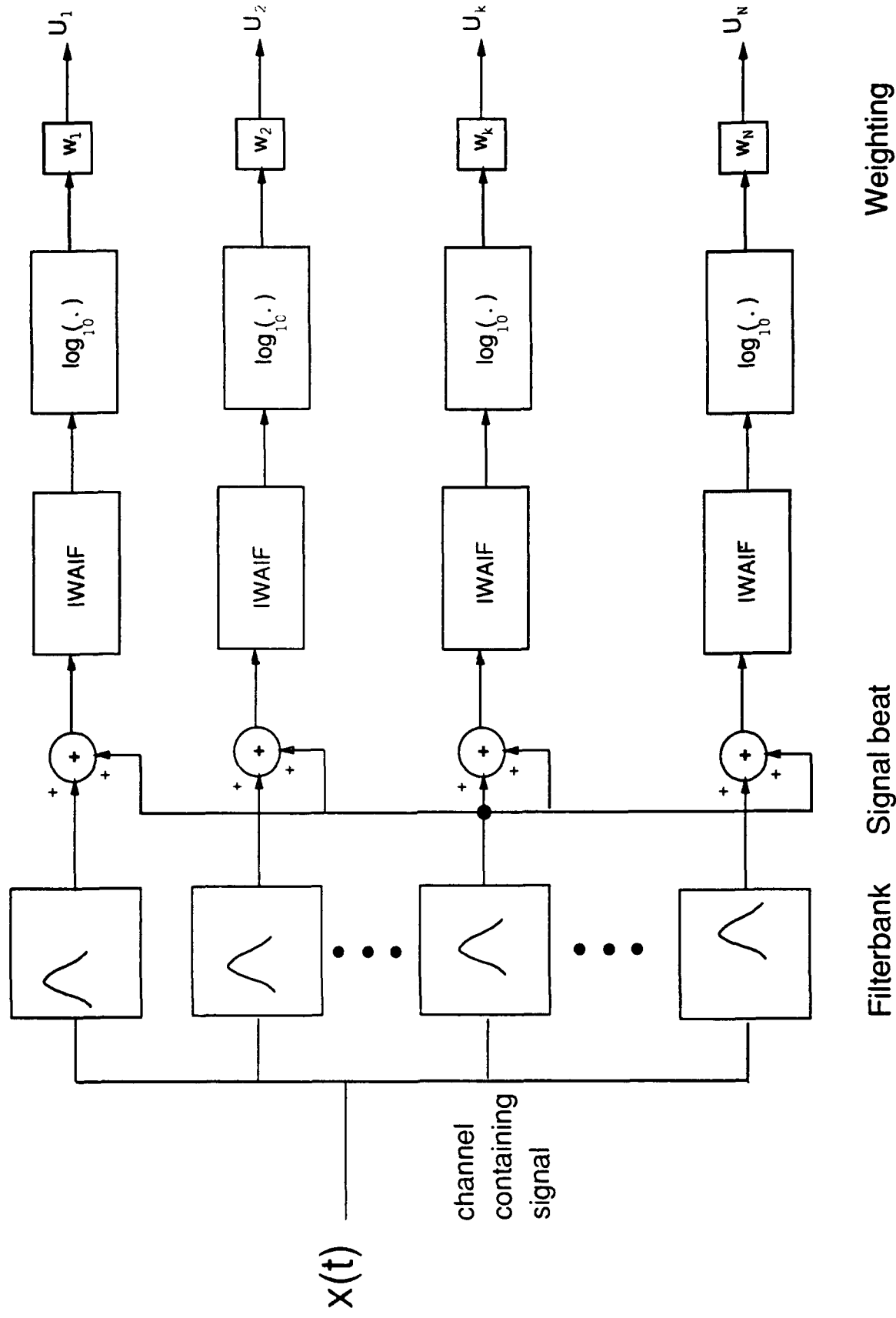


Wideband Model

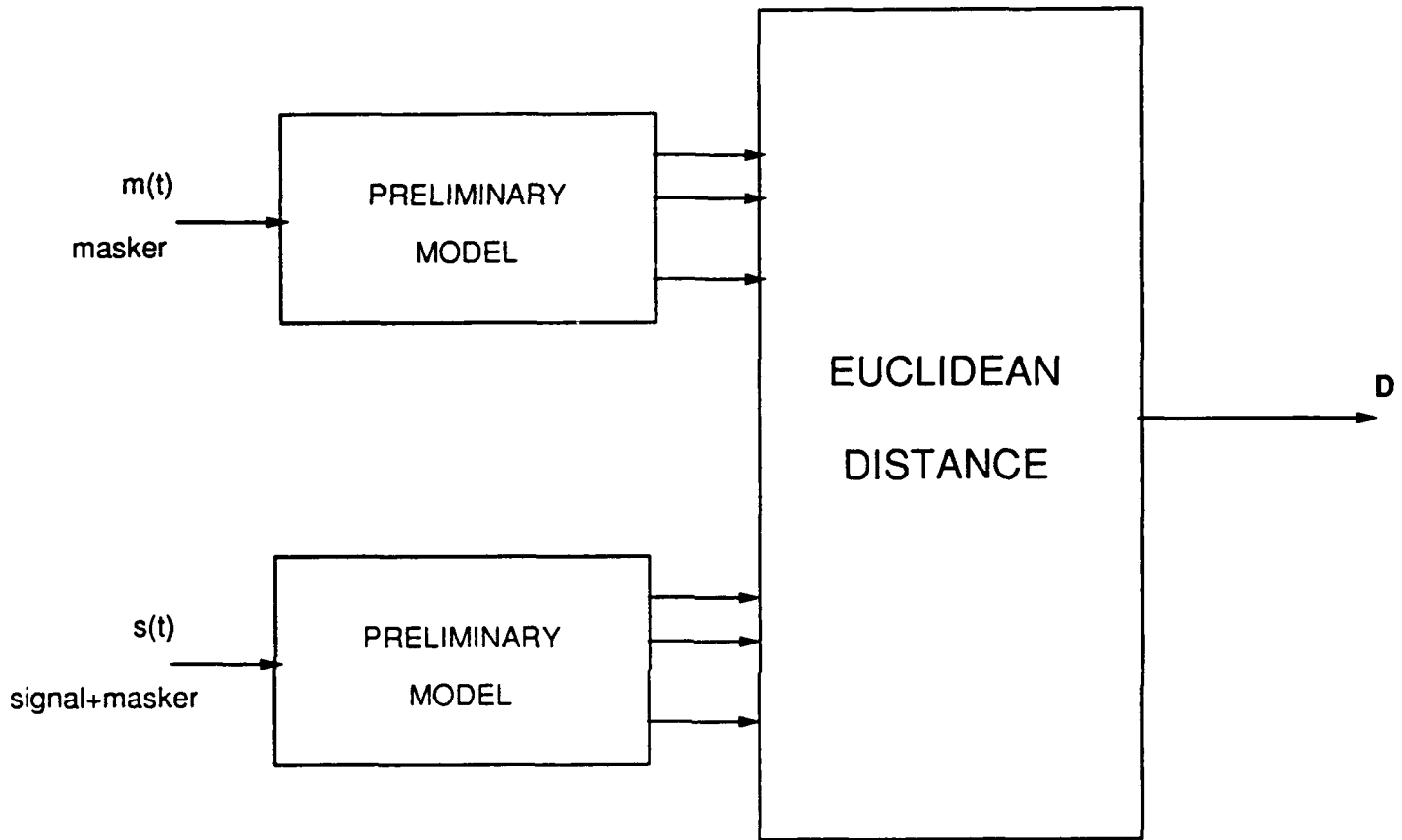


Filterbank Model

Preliminary Model



A Distance Measure



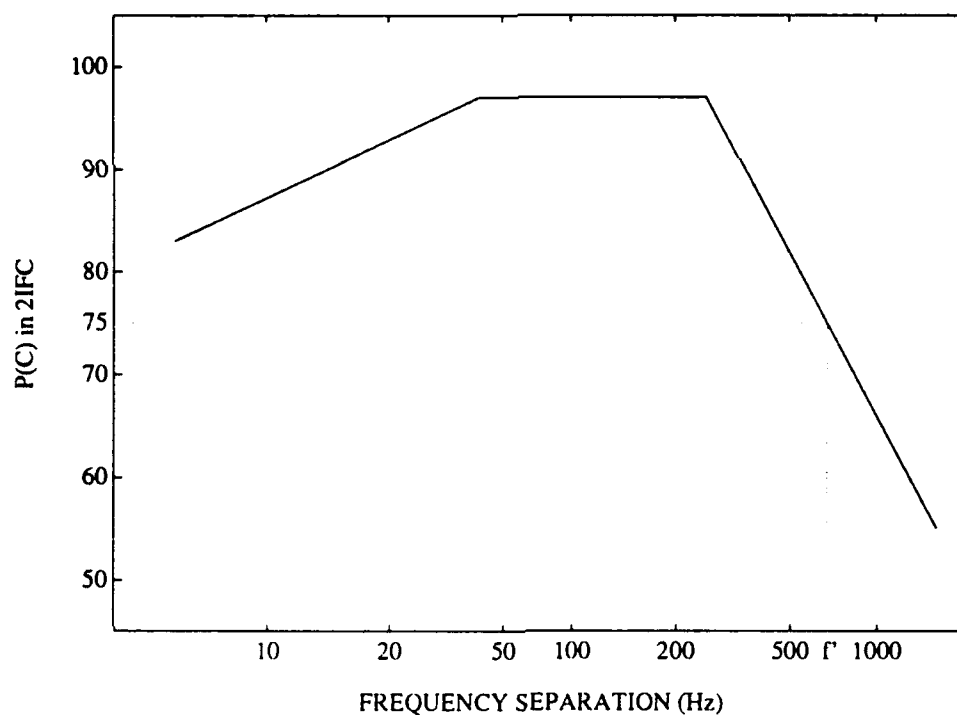
Computation of the distance measure, D ,
between $s(t)$ and $m(t)$

Distance measure,

$$D^2 = \sum_{i=1}^N \left[w_i \log \frac{IW_{s_i}}{IW_{m_i}} \right]^2$$

s_i : Output of channel i beat with the signal channel for $s(t)$

Application to two component complex tones

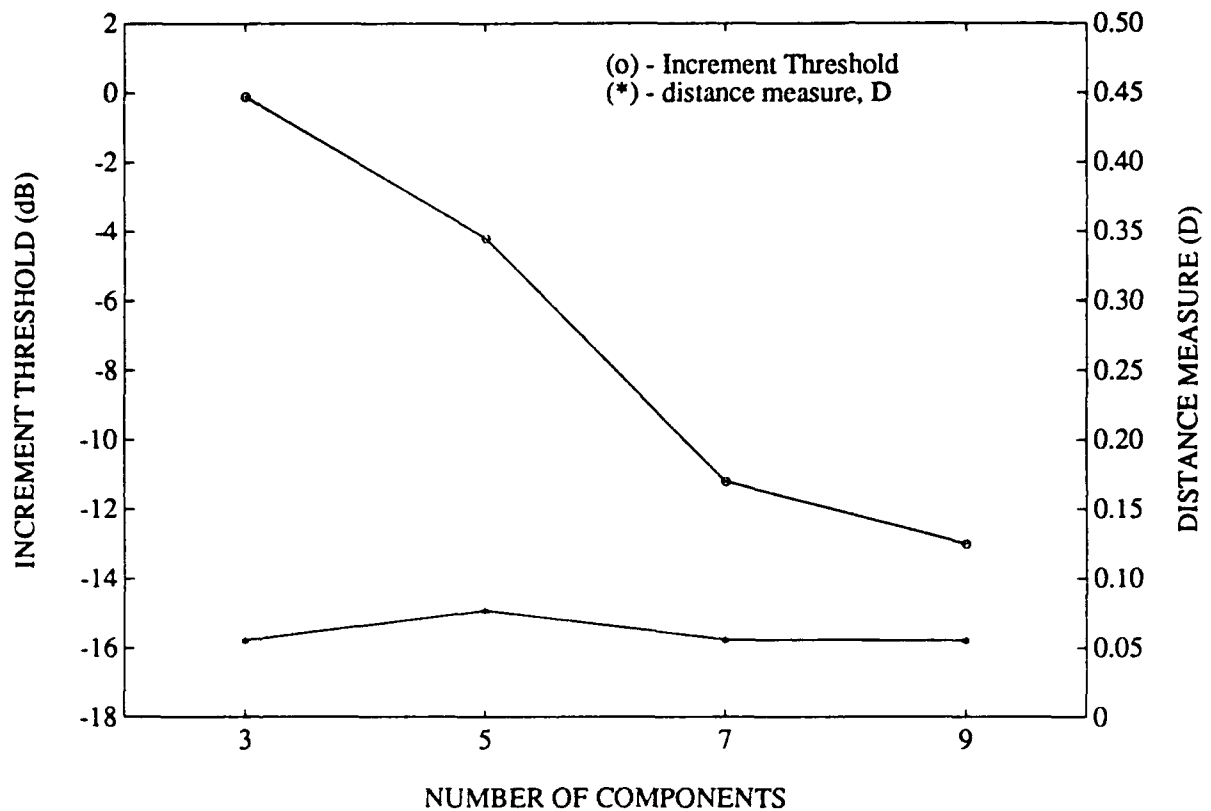


f' : -75% estimate

Distance measure, D , for two component complex tones (Feth & O'Malley 1977)

Center frequency in Hz.	-75% estimate in Hz.	Distance measure D
250	223	0.0635
500	364	0.0543
1000	692	0.0619

Application to Auditory Profile Signals



Distance measure, D , for profile signals
(Green, Mason, & Kidd 1984)

- D varies very little indicating equal listener performance.

Conclusions

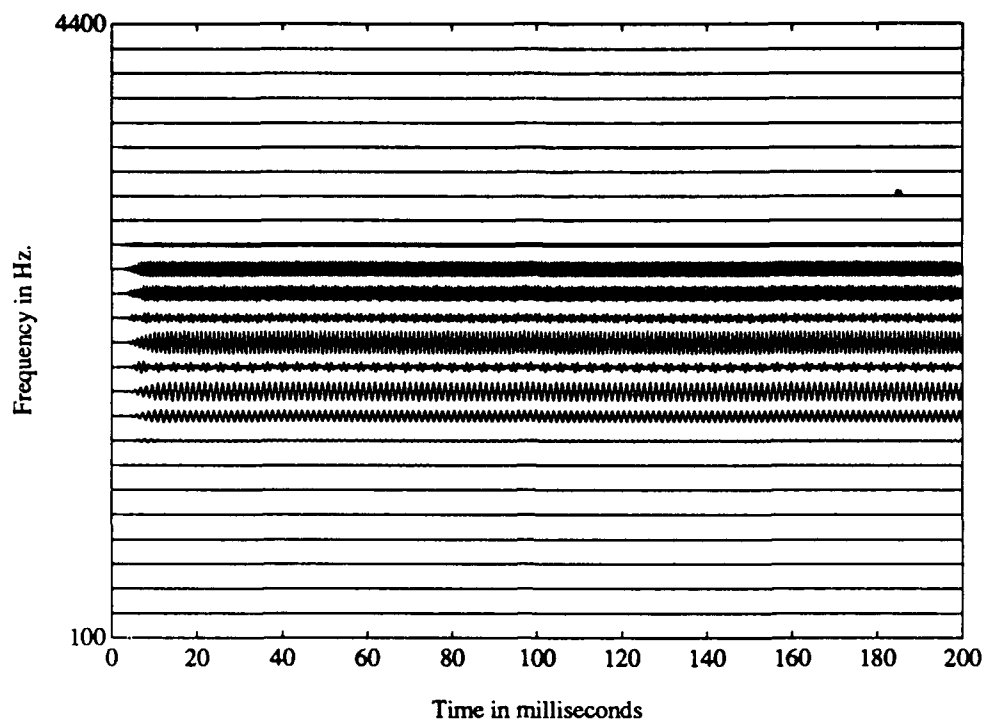
- Extension of EWAIF theory to wideband signals after accounting for the auditory filterbank.
- Definition of a distance measure between two signals to be discriminated.
- Results indicate the distance measure accurately reflects listener performance in discriminatory tasks involving profile signals and two component complex tones.
- Future applications to comodulation masking release, modulation masking and speech recognition.

References

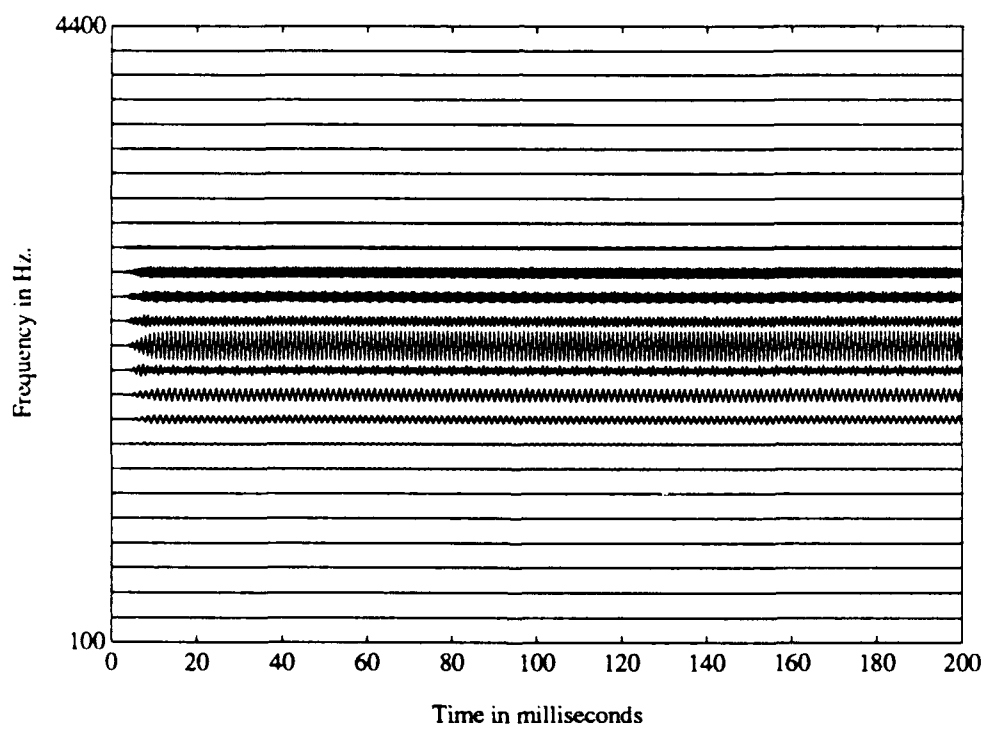
1. Feth, L. L. (1974) "Frequency discrimination of complex periodic tones," *Percept. Psychophys.*, 15, 375-379.
2. Feth, L. L. and O'Malley, H. (1977) "Two-tone auditory spectral resolution," *J. Acoust. Soc. Am.*, 62, 940-947.
3. Feth, L. L., O'Malley, H. and Ramsey, J. (1982) "Pitch of unresolved, two-component complex tones," *J. Acoust. Soc. Am.*, 72, 1403-1412.
4. Feth, L. L. and Stover, L. J. (1986) "Demodulation Processes in Auditory Perception," 76-86.
5. Green, D. M. (1988) *Profile Analysis, Auditory Intensity Discrimination* (Oxford University Press, New York).
6. Green, D. M., Mason, C. R., Kidd, G. (1984) "Profile Analysis: Critical bands and duration," *J. Accoust. Soc. Am.*, 75, 1163-1167.
7. Green, D. M., Kidd, G., and Picardi, M. C. (1983) "Successive versus simultaneous comparison in auditory intensity discrimination," *J. Accoust. Soc. Am.*, 73, 639-643.
8. McGillem, C. D. (1979) "Hilbert Transforms and Analytic Signals," Unpublished notes on signal processing, Purdue University.
9. Patterson, R., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1988) "An efficient auditory filterbank based on the gammatone function," Annex C of the SVOS Final Report (Part A: The Auditory Filterbank).
10. Patterson, R., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1988) "Implementing a gammatone filterbank," Annex C of the SVOS Final Report (Part A: The Auditory Filterbank).

Acknowledgements

Work supported in part by the Air Force Office of Scientific Research and the Ohio State Research Challenge Funds.

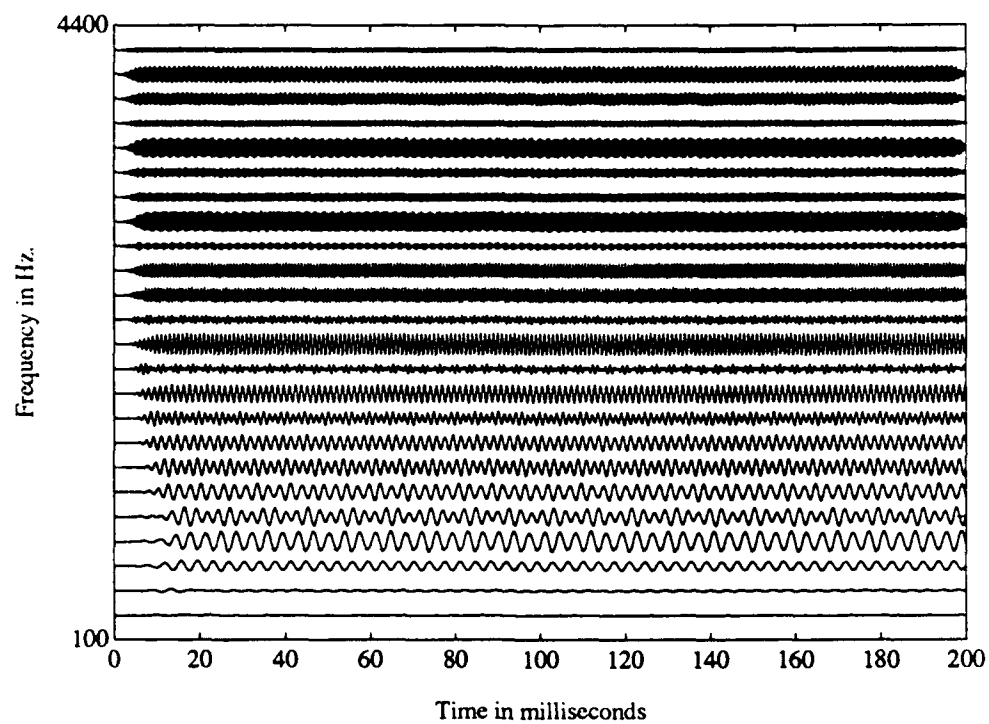


(a)

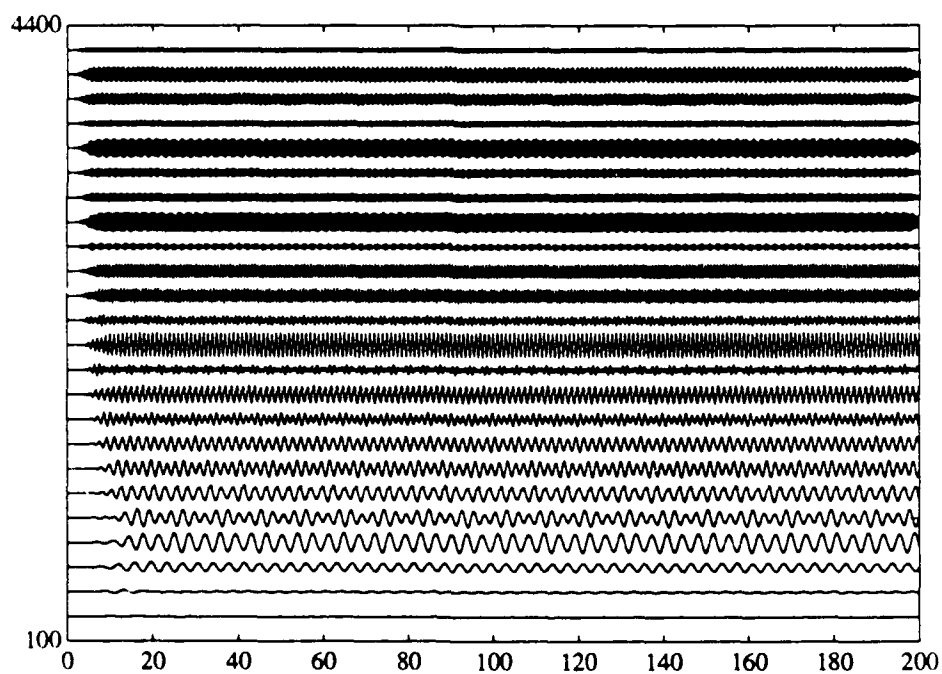


(b)

Patterson-Holdsworth gammatone filterbank output to 3-component profile signal. (a) Flat (b) Standard incremented.



(a)



(b)

Patterson-Holdsworth gammatone filterbank output to 9-component profile signal. (a) Flat (b) Standard incremented.

I2. Discrimination of frequency-glide direction.

Mary E. Neill and Lawrence L. Feth,
Speech and Hearing Science,

Ohio State University, Columbus, Ohio

ABSTRACT

It is often assumed that the relevant information in formant transitions includes the direction, as well as, the extent of frequency excursions. We have devised a study to determine the ability of listeners with normal hearing to determine the direction of linear frequency modulation of sinusoidal tones. To more closely approximate the listener's task in processing speech, music or other environmentally-important sounds, the initial frequency of each transition was selected at random from within a pre-defined range. Thus, for each interval of a 2Q, 2AFC listening task, the listener hears a frequency glide of the same duration and extent. In one of the two middle intervals, the direction of the transition is reversed to produce the "signal". Since each of the four glides begins and ends on frequencies selected at random, the listener cannot rely on simple pitch differences to determine which interval contained the reversed glide. Our preliminary results will be discussed with emphasis on the effect of the width of the "roving frequency" range on listener performance. (Work supported by a grant from AFOSR.)

INTRODUCTION

Speech information is conveyed by the dynamics of formant transitions. The extent and direction of formant transitions, especially the second formant, provide acoustic cues useful in the identification of place of articulation of consonants. Previous studies using frequency-modulated (FM) sinusoidal tones with normal-hearing listeners, indicates that the listener may use pitch cues derived from differing endpoint frequencies to determine the direction of linear frequency transitions (Nabelek, Nabelek and Hirsh, 1970; Carlyon and Stubbs, 1989). We attempted to render these pitch cues unreliable by randomizing the starting frequency of each of the glide tones presented in a two-cue, two alternative forced choice (2Q,2AFC) task. Our procedure is a frequency domain analog of the roving level paradigm used in the study of intensity perception (Berliner and Durlach, 1973) and in profile analysis studies (Green, 1988).

The preliminary work described here was designed to determine the boundaries of the range over which to randomize starting frequency. To determine the range of starting frequencies required, listeners were asked to distinguish between two linear FM glides differing only in the direction of their transitions. Initially, the starting frequency (or ending frequency, for falling glides) was fixed at 1000 Hz. Listener performance was assessed for small frequency excursions. The task was simply to indicate which of the two center listening intervals contained a glide falling in frequency when the remaining three intervals contained glides rising in frequency. Once this base-line performance was established, the starting frequency was selected from a uniform rectangular distribution with a mean of 1000 Hz. With this "roving frequency" procedure, the extent of frequency excursions was increased to again produce psychometric functions.

PROCEDURES

SUBJECTS

Five listeners with normal hearing (thresholds better than 15 dB HL re ANSI 1969) from 250- to 8000 Hz participated in the study. The age of the listeners ranged from 20 to 24 years. Prior to data collection, listeners were well-practiced at the task; all except S1 had previous experience in psychoacoustic experiments. Subjects #1 and #3 have had some musical training.

SIGNALS

Sinusoids with linear frequency modulation were generated on-line using an Ariel DSP-16 signal processing board mounted in a Zenith 159 microcomputer. All signals were generated at a 100 kHz sampling rate, and low-pass filtered at 8 kHz. The frequency sweep for each glide in a block of 50 trials was fixed, but the starting frequency could be selected from a uniform rectangular distribution centered on 1000 Hz. Signal duration was always 60 msec. including 5-msec. rise and decay times. Signals were either rising (UP) or falling (DN) linear frequency glides. In the roving frequency conditions, the starting frequency for each UP glide was selected at random from within a pre-defined range; DN glides ended on the selected frequency. The widths of the frequency range examined were 50-, 100-, 200- and 400 Hz., centered on 1000 Hz. Frequency excursions for the glides were 50-, 20-, 15-, 10-, 5-, and 2 Hz.

METHODS

Listeners were seated in separate sound-isolated booths facing a monitor and a color computer used for entering their responses. Three subjects could listen simultaneously. Signals were presented at 50 dB SL through one side of a Sennheiser HD414SL headset. For each interval of the 2Q,2AFC task, the listener heard a frequency glide. The target was a DN glide which appeared only in one of the two middle listening intervals (interval two or three). Correct response feedback was given after each trial. Data were collected in blocks of 50 trials. All data points shown were derived from the average of at least six blocks, that is 300 trials.

Figure 1 illustrates the listening task for the fixed and two "typical" roving starting frequency conditions. In the top row, each FM glide begins, or ends, at 1000 Hz. The target is shown in interval three. For the "roving" starting frequency conditions, the glide in each listening interval began on a different frequency. In the second row, the range of starting frequencies covers 50 Hz, centered on 1000 Hz. The sweeps shown extend beyond the width of the 50-Hz range. The target is shown in interval two. In the bottom row, the range of starting frequencies covers 400 Hz, much wider than any sweep width used in the study. The target is shown in interval three.

RESULTS

Psychometric functions for each of the five listeners are shown in Figure 2. Percent of correct discriminations is plotted as a function of the extent of frequency sweep of the glide. Each panel shows results for the fixed condition and for the four random starting frequency ranges. The 75% correct discrimination point is called the direction discrimination threshold. For the fixed condition, listeners are able to discriminate UP from DN glides in a 5 Hz frequency sweep of 50 msec. duration, with the exception of S4, who requires approximately 15 Hz frequency sweep. Introducing a random starting frequency degrades the discrimination performance for every listener; however, as the width of the range of starting frequency increases, the psychometric functions remain parallel for each listener.

For S1, the direction discrimination threshold moves from less than 5 Hz in the fixed condition, to about 25 Hz when the range of starting frequency reaches or exceeds 100 Hz. S1s psychometric functions coincide for the frequency ranges of 100 Hz, 200 Hz, and 400 Hz. Performance for S2b and S3 is similar to S1, however S3s direction discrimination threshold is around 30 Hz. Most of the listeners exhibited a relatively long learning period. Except for S1 (who had not participated in prior psychoacoustic experiments), each demonstrated poor direction discrimination thresholds in both fixed and random starting frequency conditions. After extensive practice, results for S2 and S3 resemble those for S1. To illustrate this, early performance of S2a, resembles that of S4, where the 75% correct point for the fixed condition lies just above 10 Hz, and thresholds for the random starting frequency conditions range from 15 to 100 Hz. Listeners S4 and S5 have just begun to participate in the experiment. We expect that with more practice, their performance will be similar to the initial three listeners.

CONCLUSIONS

For the preliminary data shown here we find that:

- Randomizing the starting frequency of linear FM glides increases the glide direction threshold approximately five-fold.
- Increasing the range of the random starting frequency beyond 100 Hz around 1000 Hz has no effect on performance.
- Most listeners require extensive practice to reach asymptotic performance.

REFERENCES

- Berliner, J.E. and Durlach, N.I. (1973). Intensity Perception. IV. Resolution in a roving-level discrimination. *J. Acoust. Soc. Amer.* 53, 1270-1287.
- Carlyon, R.P. and Stubbs, R.J. (1989). Detecting single-cycle frequency modulation imposed on sinusoids, harmonic and inharmonic carriers. *J. Acoust. Soc. Amer.* 85, 2563-2574.
- Green, D.M. (1988). *Profile Analysis: Auditory Intensity Discrimination* New York: Oxford University Press, Inc.
- Nabelek, I.V., Nabelek, A.K. and Hirsh, I.J. (1970). Pitch of tone bursts of changing frequency. *J. Acoust. Soc. Amer.* 48, 536-553.

FIGURE 1. Glide Direction Discrimination Paradigm.

Top - fixed starting frequency. Middle - random starting frequency shown with range less than the sweep width. Bottom - random starting frequency with range much larger than the sweep width.

FIGURE 2. Results for Glide Direction Discrimination.

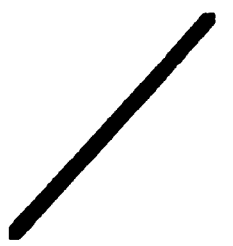
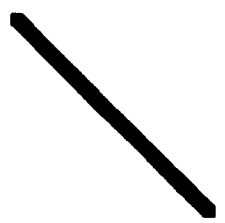
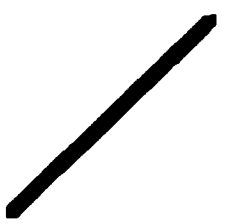
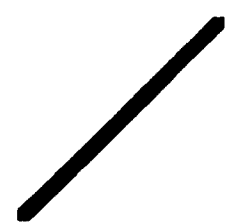
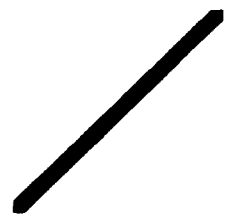
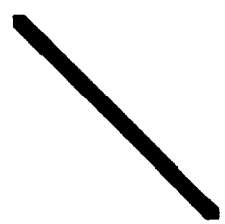
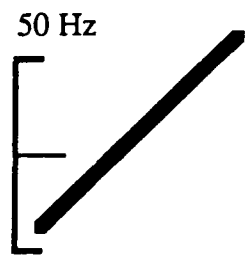
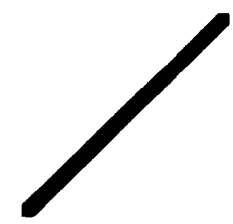
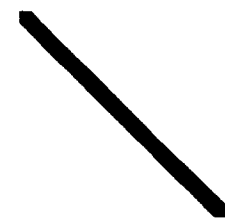
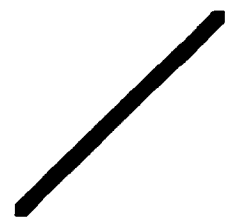
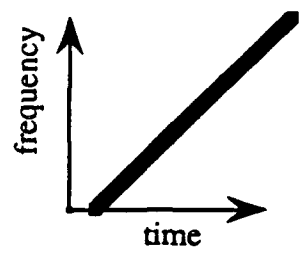
Psychometric functions for five listeners with normal hearing. Percent correct discrimination as a function of frequency sweep width in 2Q, 2AFC. Results for listener 2 are shown for early performance (S2a) and after extensive practice (S2b).

1

2

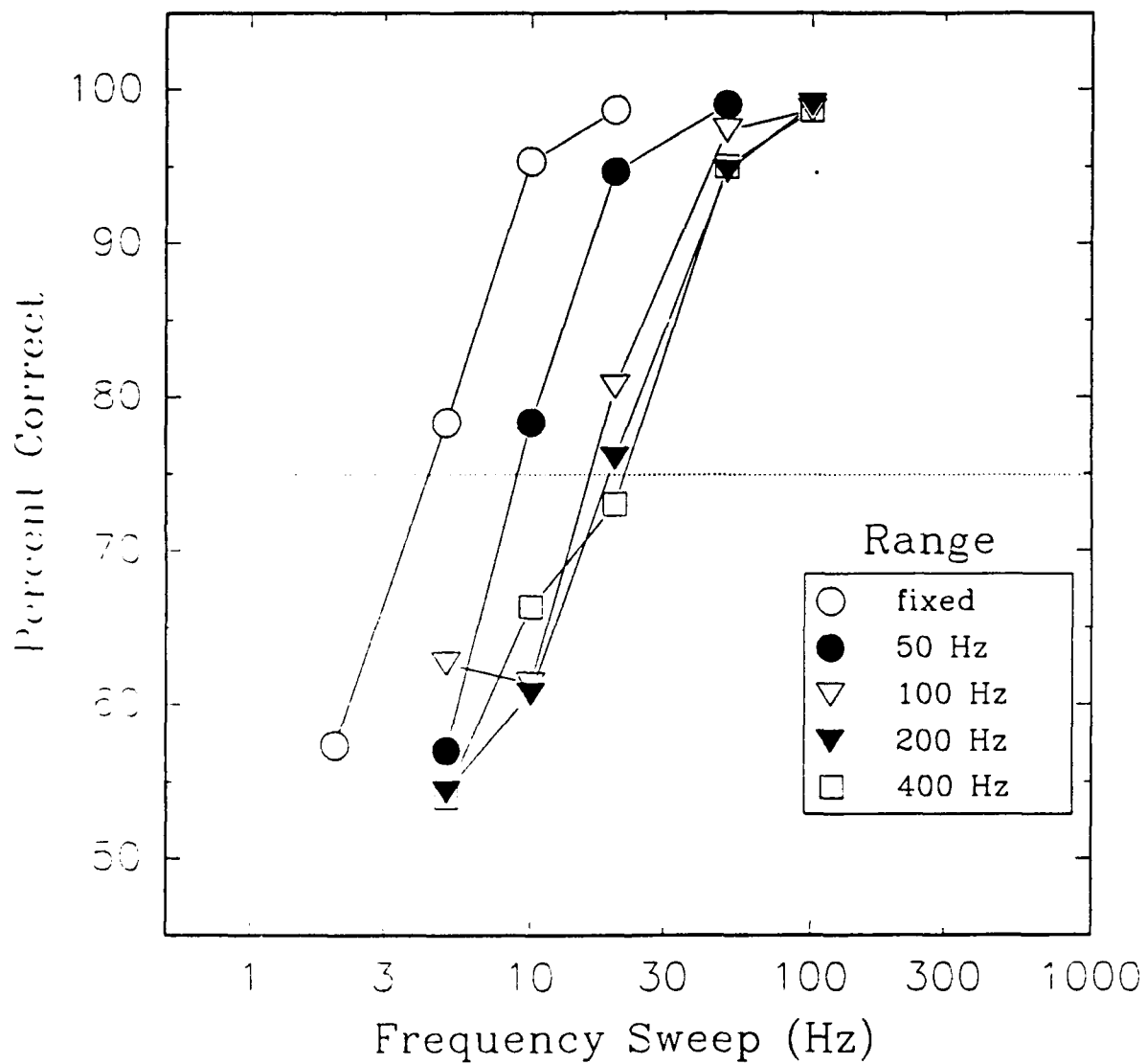
3

4

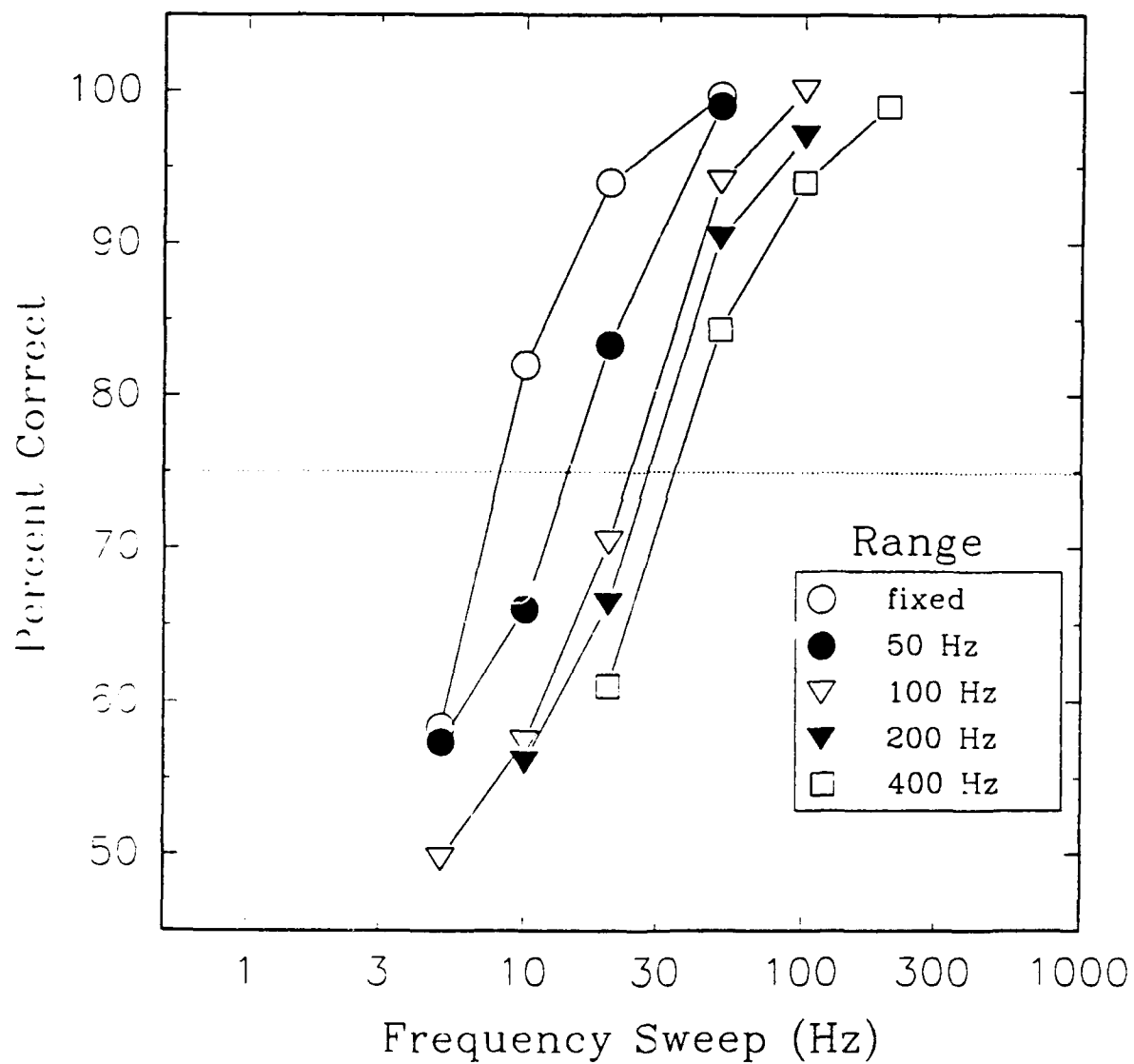


GLIDE DIRECTION DISCRIMINATION

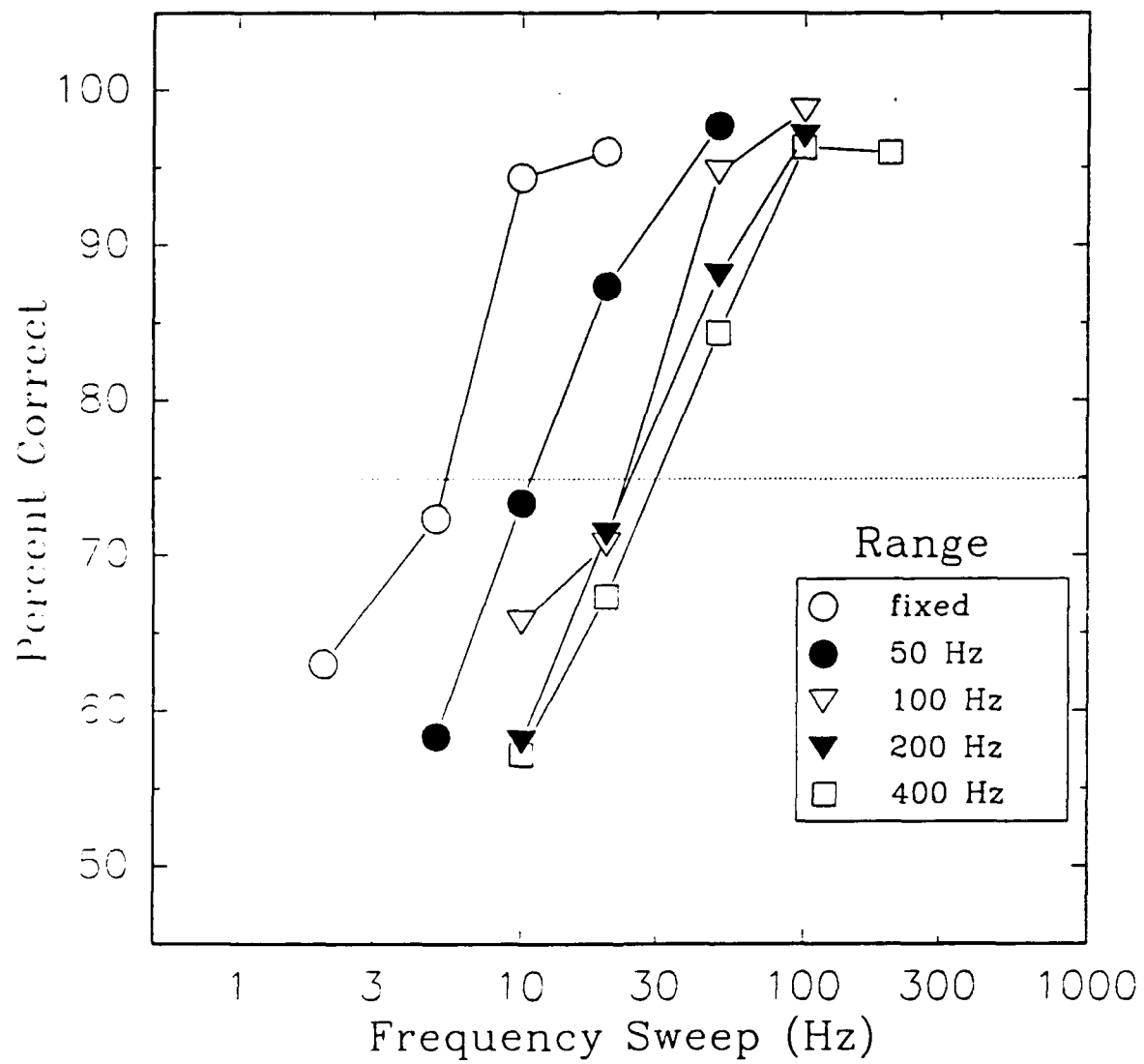
Glide Direction Discrimination S#1



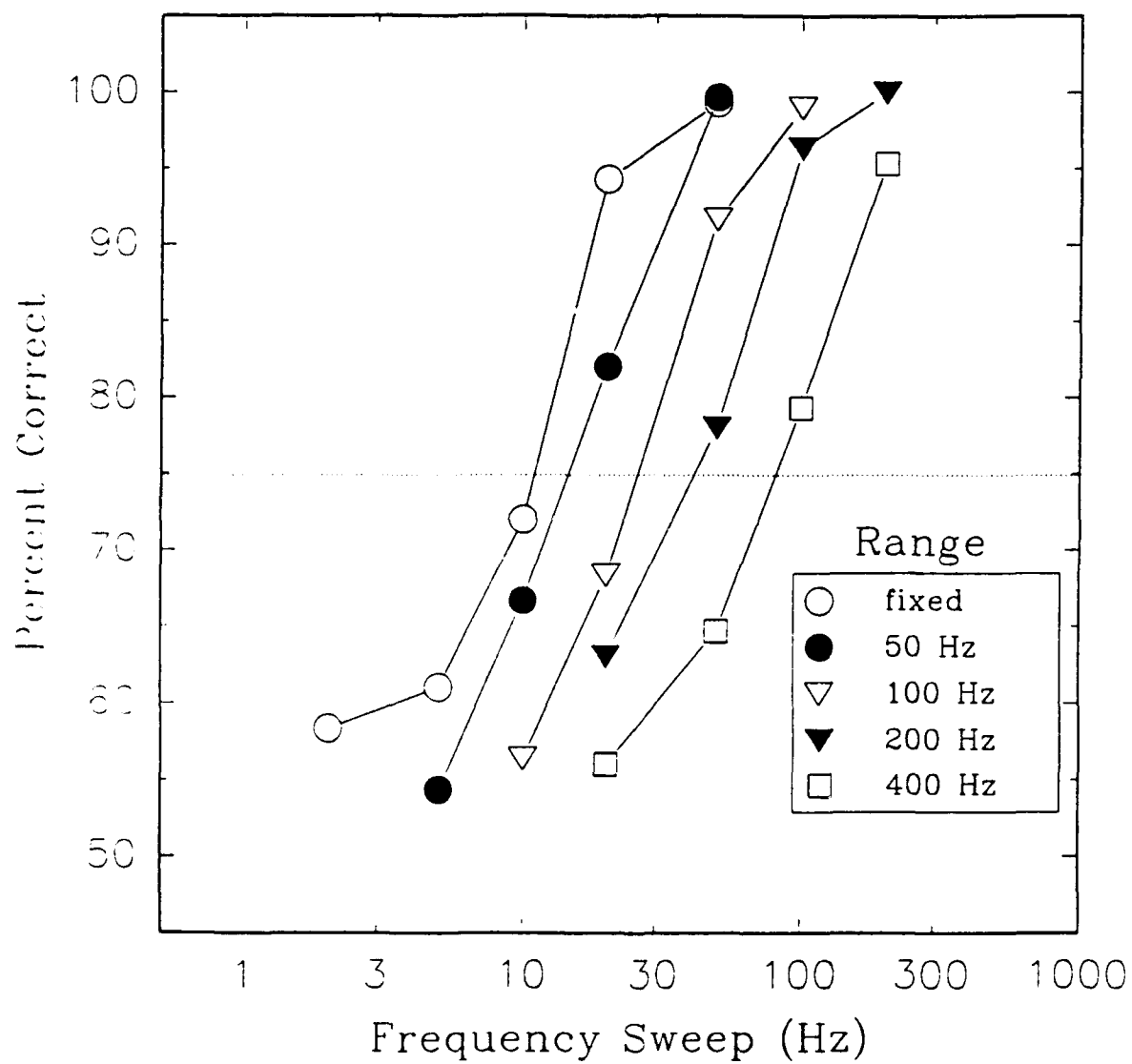
Glide Direction Discrimination S#2b



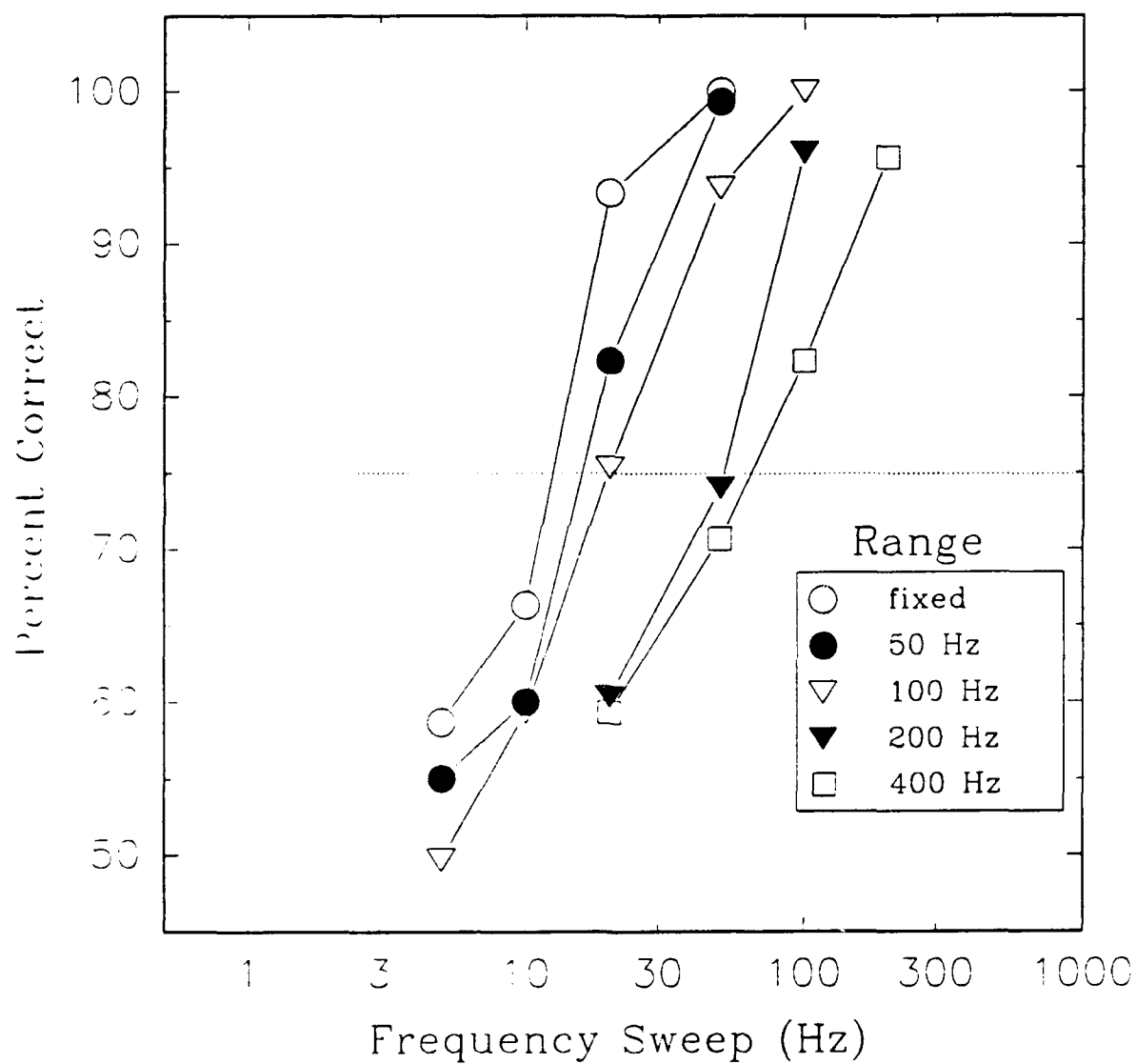
Glide Direction Discrimination S#3



Glide Direction Discrimination S#4



Glide Direction Discrimination S#2a



Glide Direction Discrimination S#5

